SCIENTIFIC REPORT 2022

idiap RESEARCH INSTITUTE

Contents

List of Figures

1 Executive summary

Statutes and areas of activities: Founded in 1991 by the State of Valais, the City of Martigny, the Ecole Polytechnique Fédérale de Lausanne (EPFL), the University of Geneva, and Swisscom, the Idiap Research Institute[1] is an independent, non-profit, *Research Foundation* devoted to advanced research, training, developments and technology transfer in the areas of *Artificial and Cognitive Intelligence*, and is among the world-level leaders in the field.

Based in Martigny (Valais, Switzerland), the Institute is particularly active in advanced signal processing, machine learning, and knowledge management, applied to multiple domains, including voice and visual recognition, multimedia data mining, human-machine interaction, robotics, language analysis, and bio-imaging. Idiap is involved in numerous research projects at local, national and international levels. As part of its core missions, Idiap is also very active in academic training (Masters and PhD students), as well as in technology transfer, through collaboration with various industries, or through direct spin-off (and its incubator IdeArk S.A.[2]).

Affiliation and budget: Since its inception, Idiap has been a completely independent research institution, although academically affiliated with EPFL and the University of Geneva. Since July 2008, Idiap and EPFL are agreeing on a "Joint Development Plan", which is revisited and signed every 4 years. The last version was signed June 26, 2021, applicable until December 2024. Considered as part of the "ETH Strategic Domain" (Research Institution of National Importance[3]), Idiap is accredited and co-funded by the Federal Government, the State of Valais, and the City of Martigny, for a total averaging 45-50% of its annual budget, the remainder coming from competitive projects and industry. With an initial budget in 1991 of around 500 KCHF, the Idiap budget has been steadily growing over the last 30 years to be today well above 14.5 MCHF/year.

Projects: Idiap is active in numerous national and international projects (with an average of around 50 projects active simultaneously), while also being active in technology transfer and research contracts with industry. From 2001 to 2013, Idiap was also the Host Institution of one of the major National Centres of Competence in Research (NCCR) on "Interactive Multimodal Information Management" (IM2[4]). Project distributions and dynamics, including statistics of acceptance rates across all funding instruments, are presented in Section 5, page 53.

Staff and publications: With 16 research groups active in 2022 (Section 2.3.2, page 6) and a staff of approximately 150 people mainly composed of senior researchers, R&D engineers, post-docs, and PhD students, Idiap regularly generates a large amount of high-quality peer-reviewed international scientific publications per year (in addition to an equal amount of internal Research Reports), all available from a dedicated web site[5], automatically synchronised with EPFL Infoscience[6].

As part of our 2021-2024 Research Plan, "**Cross Research Groups**" (CRG) started to be created to foster collaborations between Idiap Research Groups. This new kind of CRG research structure was laid down a few years ago by Idiap's management with the objective to reinforce the institute's ability to foster collaborations, really critical in increasing technology transfer and impact on society. Their aim is to have long-term impacts on society thanks to an approach mixing both business oriented solutions and interdisciplinary scientific groundwork. In 2022, we initiated three CRGs (as part of the 16 Research Groups): (1) Neuro-Symbolic Learning & Reasoning (led by Dr. Andre Freitas), (2) AI for Trust (lead by Prof. Sébastien Marcel, and (3) Human-Centered Robotics & AI (led by Dr. Emmanuel Senft).

[1] www.idiap.ch
[2] www.theark.ch
[3] https://www.sbfi.admin.ch
[4] www.im2.ch
[5] publications.idiap.ch
[6] https://infoscience.epfl.ch/

Open-source software libraries: Idiap pursues an active policy towards the open-source release of high-quality software libraries, after having been cleared through a strict quality control process and clear IPR and distribution rules. Idiap's web site[7] refers to 133 open-source libraries.

Public datasets: Idiap has always been driven towards open data and researchers are encouraged to have their data published through a FAIR repository [8] to promote reproductible science. All data distribution are centrally managed and comply with SNSF guidelines on open data as well as Swiss and European regulations on data protection (FADP and GDPR). Idiap web [9] site is currently referencing 80 publicly available datasets.

Reproducible research: All our work is always driven by fully reproducible frameworks[10] and Idiap regularly releases open source codes accompanying its publications[11].

Quality indicators: To calibrate and optimise its performance, Idiap regularly screens the progress of the Institute against multiple indicators related to its key missions, i.e., research, technology transfer, training, and competitive funding. Thus, in addition to the present Research Report, our 2022 Self-Assessment Report, provides a concise and factual picture of where we stand in terms of our organisation, group structure, human resources, quality of projects and research staff, academic and professional activities, publications, teaching and technology transfer activities.

Technology transfer is also one of Idiap's key missions. In our separate 2022 Self-Assessment report, we are reporting on the excellent performance and continuous increase of our technology transfer activities, patenting, collaboration with industries, also attracting startups, as well as initiating our own spin-off companies. In the present report, Section 7, page 79, we list the patents currently in our portfolio.

[7] https://www.idiap.ch/en/scientific-research/resources
[8] https://www.snf.ch/en/dMILj9t4LNk8NwyR/topic/open-research-data
[9] https://www.idiap.ch/en/scientific-research/resources
[10] See, e.g., https://gitlab.idiap.ch/bob/bob.ip.binseg
[11] See, e.g., https://www.idiap.ch/software/pbdlib/

2 Idiap Structure and Research Areas

2.1 Research Areas

Idiap's research activities span five broad thematic areas, all directly related to our central themes and competencies of **Signal Processing**, **Machine Learning**, and **Social Intelligence for Society**: perceptual and cognitive systems, human and social behavior, multimodal/multimedia information processing and presentation interfaces, biometrics security and privacy, and machine learning, which are summarized, with a set of descriptive keywords in Table 1 below.

In addition to these transverse research themes, new activities have been recently developed, towards biomedical applications (through the groups in Computational Bioimaging and Genomics) and environmental modeling.

Idiap maintains a policy of continuous growth and diversification in new and complementary research directions, while building upon its long-standing institutional theme of "Human and Media Computing".

Research Areas	***Keywords***
Perceptual and cognitive systems	*Speech and audio processing, computer vision, document processing, robotics, natural language processing, machine translation, computational cognitive science*
Human and social behavior	*Social media, verbal and nonverbal communication analysis, smartphone sensing, computational social science*
Information and presentation interfaces	*Multimedia information systems, user interfaces, personalization, system evaluation, mobile HCI using big data, data driven services*
Biometrics Security and Privacy	*Face recognition, speaker recognition, vein recognition, gait recognition, multimodal fusion, soft-biometrics, remote photoplethysmography, presentation attack detection (anti-spoofing), morphing attack detection, deepfakes detection, template protection, privacy preservation, mobile and wearable biometrics*
Machine learning	*Statistical and neural network based machine learning, continual learning, learning over vision and language, robust learning and handling of dataset biases, responsible AI*

Table 1: Overview of Idiap research areas. These research areas are currently covered by twelve research groups, listed in Section 2.3 below, and are designed to be as pro-active as possible in covering the applications areas presented in Figure 2.

2.2 Application Areas

While Idiap's research areas (Table 1) are quite generic and have a strong transverse nature, those are also key enablers of multiple applications. The key application areas targeted at Idiap are summarized in Table 2 below. All of those application areas have a strong cross-research group nature, hence the creation of "Cross Research Groups" (CRG) discussed in Section 2.3.3

Application Areas	***Keywords***
Human-human & human-machine interaction	*Voice and gesture controlled devices and robots, hand-free control, spoken language systems, translation systems, social robotics, user profiling*
Exploitation of rich multimedia archives (audio, video, text)	*Semantic indexing, knowledge graph, object detection and recognition, audio-video content filtering (summarization and recommendation), broadcast data analysis, scanned document analysis, analysis of cultural heritage media*
Collaborative and creative systems	*Remote meeting assistance, smart meeting room, video-conferencing, multimedia indexing and access, cross-lingual collaboration, interaction analysis, dynamics of negotiation*
Healthcare and bio-engineering	*Smart management of patient data, prosthesis (hearing aids, artificial body parts, voice banking), bio-systems modeling, inclusive ICT, bio-medical document and data processing.*
Entertainment	*Multilingual gaming, remote family games, togetherness*
Mobile computing	*Signal processing for mobile platforms, mobile social networks, participatory sensing*
Security and risk management	*Biometric security, deepfakes, access control, mobile biometry, multi-sensor fusion, speaker identification, video monitoring of areas/activities, natural risk modeling, intrusion detection, crowd management*
Home automation (domotics)	*Multi-sensor activity analysis, adaptation to users' behavior, efficient use of energy, home safety and security*
Energy	*Energy grids, multiple sensor and smart meter networks, large-scale sensor data integration, modeling of behaviors to anticipate demand, safer, cheaper, and cleaner energy production*
Smart processes	*Industry 4.0, smart manufacturing, predictive maintenance, fleet management, additive manufacturing, capture and management of industrial data.*
Smart cities	*Ecology, environment management, reduction in pollution, traffic and noise, better use of roads*

Table 2: Idiap application areas with several examples for each of them.

2.3 Internal Structure

2.3.1 Overview

The main research themes, introduced in Section 2.1, are currently covered by 14 Research Groups (RG), and 4 (under development) Cross-Research Groups (CRG), presented in the Section 2.3.2. The RG and CRG leaders, explicitely named, are all *Principal Investigators (PI)* of research projects.

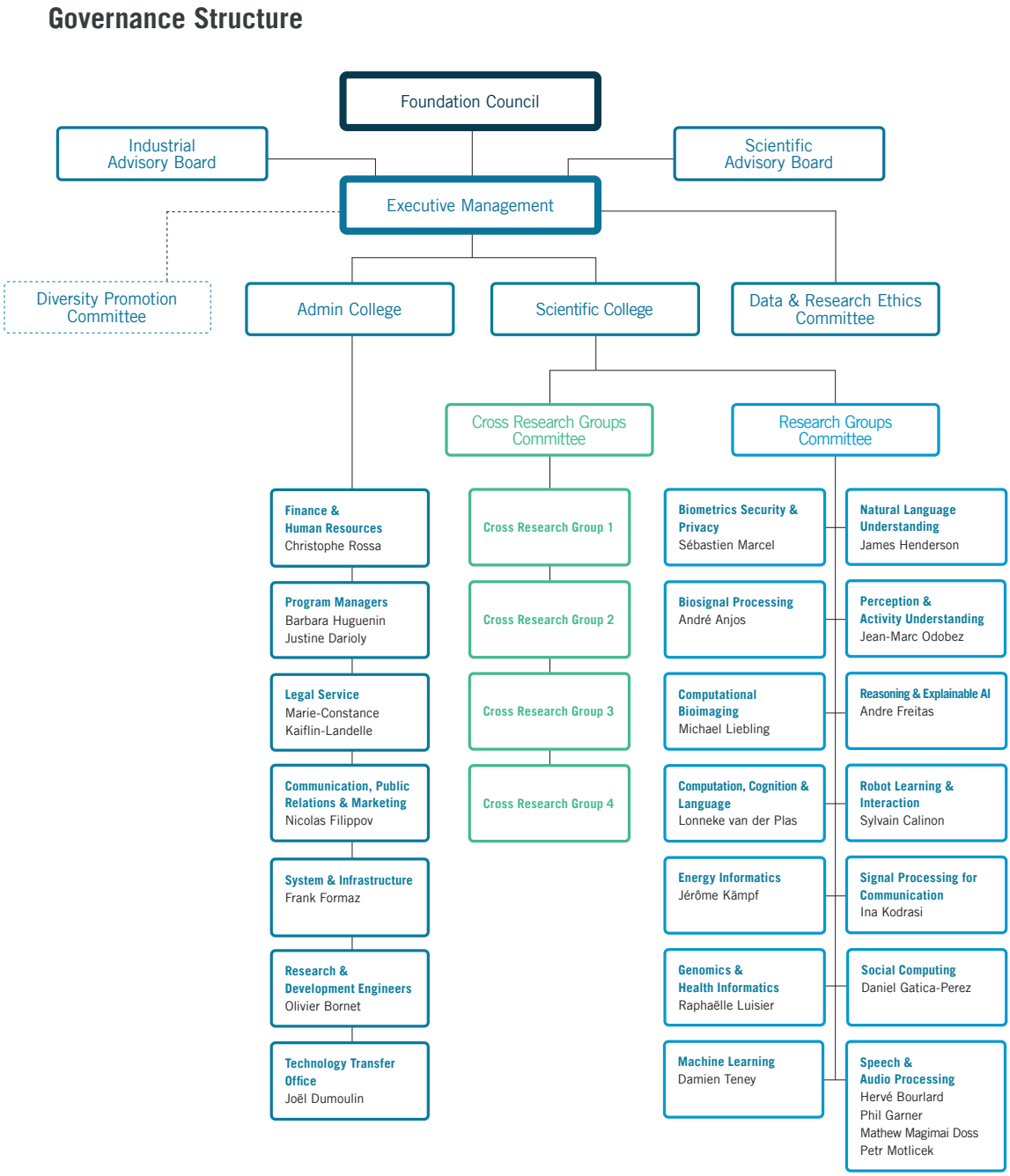


Figure 1: Idiap management and operational structure, including the 14 Research Groups on the right side, the (under-development) Cross Research Groups (CRG) at the middle, as well as the administration and services on the left side of the organigram.

2.3.2 Research Groups (RG)

In 2022, the research areas presented in Figure 1 were covered by twelve research groups. The activity of each research group (from left to right branch in Figure 1) can be summarized as follows.

1. **Speech and Audio Processing (Prof. Hervé Bourlard, Dr. Philip N. Garner, Dr. Mathew Magimai-Doss, Dr. Petr Motlicek)**

   Speech processing has been one of the mainstays of Idiap's research portfolio for many years. Today it is still the largest group within the institute, and Idiap continues to be recognized as a leading proponent in the field. The expertise of the group encompasses multilingual speech recognition (with applications beyond the speech), multilingual text-to-speech conversion, and generic audio processing – covering sound source localization, microphone arrays, speaker segmentation and diarization, audio indexing, speech coding (e.g. for low bit-rate scenarios), perceptual background noise analysis for telecommunication systems and speech signal processing for clinical applications.

   See Section 3.1, page 11, for the 2022 progress report.

2. **Social Computing (Prof. Daniel Gatica-Perez)**

   The social computing group studies how people and technology interface in everyday life. The group's research combines theories and methods from computing (ubiquitous computing, social media, machine learning) and the social sciences to analyze human and social behavior and design systems that support individuals and communities. Current lines of research include included mobile crowdsensing for health and cities; human-centered analysis of social media and news; and privacy-preserving machine learning.

   See Section 3.2, page 17, for the 2022 progress report.

3. **Machine Learning (Prof. François Fleuret)**

   The goal of the Machine Learning group is the development of new techniques based mainly on statistical learning, applicable to a wide range of data types and downstream applications. These applications encompass real-world signals such as images, videos, text, or lower-level signals from industrial sensors. The developed techniques aim to address challenges of real-world data such as noise, temporal drift, and distribution shifts.

   See Section 3.3, page 19, for the 2022 progress report.

4. **Perception and Activity Understanding (Dr. Jean-Marc Odobez)**

   The group aims to develop novel computational learning methods anchored in machine learning, computer vision, multimodal signal processing, or social sciences, to improve the representation and understanding of activities from real-world signals, with an emphasis on those related to humans. In particular, we investigate unsupervised learning and weakly supervised learning, user adaptation, co-training, knowledge distillation and multi-task learning, for tasks such as pose estimation, speech detection and localisation, or the recognition and analysis of non-verbal behaviors, gestures, activities or social relationships.

   See Section 3.4, page 21, for the 2022 progress report.

5. **Genomics and Health Informatics Group (Dr. Raphaëlle Luisier)** The Genomics & Health Informatics Group was created in 2019 with the overarching goal to transform our understanding and treatment of devastating and incurable human diseases by adapting cutting-edge computational methods inspired from various signal processing fields to longitudinal biological and clinical data. The group works in close collaboration with biologists and clinicians from hospitals (CHUV, Sion Hospital), academia (Francis Crick Institute, London; Kings College, London; University College London) and industry (NIBR, Suva).

   See Section 3.5, page 23, for the 2022 progress report.

6. **Robot Learning and Interaction (Dr. Sylvain Calinon)**

   The Robot Learning and Interaction group focuses on human-centered robotics applications in which the robots can acquire new skills from only few demonstrations and interactions. It requires the development of models that can exploit the structure and geometry of the acquired data in an efficient way, the development of optimal control techniques that can exploit the learned task variations and coordination patterns, and the development of intuitive active learning interfaces.

   See Section 3.6, page 25, for the 2022 progress report.

7. **Natural Language Understanding (Dr. James Henderson)**

   The Natural Language Understanding group studies deep learning for natural language processing, focusing on learning representations of the meaning of text and attention-based models of graphs. It models summarization, abstraction (textual entailment), machine translation, knowledge extraction, syntactic structure, and semantics, among other NLP problems.

   See Section 3.7, page 27, for the 2022 progress report.

8. **Computational Bioimaging (Prof. Michael Liebling)**

   This group focuses on research in computational imaging and analysis of biomedical images. This includes developing algorithms for image deconvolution and super-resolution in optical microscopy, three-dimensional tomography reconstruction from projections, and, more generally, combining unusual sensing devices and approaches with computational methods to produce images ideally-suited for the observation and quantification of complex and live biological systems.

   See Section 3.8, page 29, for the 2022 progress report.

9. **Biometrics Security and Privacy (Prof. Sébastien Marcel)**

   The Biometrics Security and Privacy group investigates and develops novel signal processing and machine learning for face recognition (2D, 3D, and near-infrared), speaker recognition, vein recognition, gait recognition, anti-spoofing (presentation attack detection), morphing attack detection, deepfakes detection and privacy-enhancement.

   See Section 3.9, page 31, for the 2022 progress report.

10. **Biosignal Processing (Dr. André Anjos)**

    The Biosignal Processing group sits in the intersection between AI and Healthcare. It studies the use of Machine Learning for computer-aided diagnosis, prognosis, decision support, and risk analysis from a variety of bio-signals, including, but not limited to, imaging, clinical, demographical, and temporal data. The group has a special focus in reproducibility and the responsible use of AI, with an interest on fairness, explainability, and safety.

    See Section 3.10, page 35, for the 2022 progress report.

11. **Energy Informatics (Dr. Jérôme Kämpf)**

    The Energy Informatics Group studies the exploitation of state-of-the-art Information and Communication Technologies to tackle global warming and climate change challenges. The aim is to increase integration of renewable and distributed energy sources by making energy systems smarter, and to increase energy efficiency beyond what improvements at component level can achieve. This includes the simulation of energy transition pathways, renewable energy production, and energy storage in a changing climate.

    See Section 3.11, page 37, for the 2022 progress report.

12. **Computation, Cognition and Language (Dr. Lonneke van der Plas)**

    The Computation, Cognition and Language (CCL) group was newly created in 2021. The group aims to build models of language that are inspired by human cognition and linguistic theory thereby working on the frontiers of what current models are capable of.

    Its current focus lies in cross-lingual transfer of language technology tools from one language to another language in low-resource scenarios, and modelling creative processes to create technology that supports human creativity.

    See Section 3.12, page 39, for the 2022 progress report.

13. **Reasoning and Explainable AI (Dr. André Freitas)**

    Established in 2021, the Reasoning & Explainable AI (ExplAIn) group aims at developing AI models which are capable of complex, abstract and flexible inference. The group operates at the interface between neural and symbolic AI methods aiming to enable the next generation of explainable, data-efficient and safe AI models. These themes closely dialogue with the need for mitigating some of the risks for the adoption of AI models in real world scenarios.

    See Section 3.13, page 41, for the 2022 progress report.

14. **Signal Processing for Communication (Dr. Ina Kodrasi)**

    The Signal Processing for Communication group focuses on signal processing and pattern recognition approaches for acoustic communication. In particular, we develop single- and multi-channel processing and pattern recognition approaches for signal enhancement, source separation, echo suppression, pathological speech detection, and hearing support.

    See Section 3.14, page 44, for the 2022 progress report.

15. **Human-centered Robotics and AI (Dr. Emmanuel Senft)**

    The Human-centered Robotics and AI group is a cross research group that operates at the interface of humans and AI technologies. Its core research is on human-robot interaction, with a focus on assistive robotics. The group also collaborates with institutional and industrial partners to integrate research done at Idiap and support its deployment in real world applications.

    See Section 3.15, page 46, for the 2022 progress report.

2.3.3 Cross Research Groups (CRG)

In Table 2, page 4, we present the main application areas, driven by *societal needs*, that can be targeted by the research currently active at Idiap.

Over the last few years, it was also observed that it is very challenging to properly address those application areas without involving (1) multiple Research Groups (RGs) and (2) additional application-specific research. Contributing to those application domains not only requires the collaboration between several research groups, but also strong multi-disciplinarity, robust platforms to transition from development to deployment, and standardization of software interfaces.

As part of our 2021-2024 Research Plan, "**Cross Research Groups**" (CRG) started to be created to foster collaborations between Idiap Research Groups. This new kind of CRG research structure was laid down a few years ago by Idiap's management with the objective to reinforce the institute's ability to foster collaborations, really critical in increasing technology transfer and impact on society. Their aim is to have long-term impacts on society thanks to an approach mixing both business oriented solutions and interdisciplinary scientific groundwork.

In 2022, we initiated three CRGs (as part of the 15 Research Groups): (1) Neuro-Symbolic Learning & Reasoning (led by Dr. Andre Freitas), (2) AI for Trust (lead by Prof. Sébastien Marcel, and (3) Human-Centered Robotics & AI (led by Dr. Emmanuel Senft), as briefly described below:

1. **Neuro-Symbolic Learning & Reasoning (Head: Dr. Andre Freitas)**: Despite its significant progress, machine learning methods are still limited towards well-defined and isolated learning and inference tasks. In contrast, a substantial part of the industrial needs and societal challenges require analytical methods which can support AI users making sense of increasingly complex problems, with limited and heterogeneous data and leveraging existing expert knowledge.

   The vision of the Neuro-Symbolic Learning & Reasoning CRG is to build the next generation of AI models which can efficiently learn and reason over heterogeneous data, dialoguing with the new strategic aim of 'doing more with less'. The group closely collaborates with industrial and clinical partners to deliver AI models which can operate under the complexity of the real-world, with a particular emphasis in the areas of Natural Language Processing and Biomedical applications. This is achieved by developing new methods which can integrate the flexibility of machine learning to the control, abstraction, and transparency of symbolic methods.

2. **AI for Trust (Head: Prof. Sébastien Marcel)**: Trust is one of the main pillars of our modern society. Citizens, customers, industry and policy makers firmly believe in the reliability of our infrastructures, services and systems. As an example, everyday we drive a car, and we trust, it can be stopped at anytime. Engineering, among other disciplines, helped us to build this perception of trust and today cars are considered a safe mean of transportation and we trust them. Nowadays, trust is tampered because of the global consequences of energy shortages or climate change, and society is exposed to new threats (e.g. disinformation, counterfeits, trafficking, business intelligence, national security, humanitarian) putting its security at risk. Societal needs and industrial demand are growing, and are likely to grow, to address these risks, and Artificial Intelligence (AI) holds many of the solutions to these emerging threats.

3. **Human-Centered Robotics & AI (Head: Dr. Emmanuel Senft)**: The Human-Centered Robotics group focuses on developing, deploying, and evaluating technologies to address societal needs. The key goals of this group are to (1) pursue research at the interface between technology and its users and (2) collaborate with industrial partners to ensure that academic research leads to positive impacts on society.

2.3.4 Administration and Services

At the administration and services level, there are seven groups: finances and human resources; communication, public relations and marketing; system and infrastructure; technology transfer; development engineers; program managers; and legal advisor. The resulting admin organization is presented on the left branch of Figure 1, page 5. The main responsibilities of the seven admin groups can be summarized as follows:

1. **Finance and Human Resources (Christophe Rossa)**: The Human Resources Department (HR) is integrated within the financial and accounting activities of Idiap and has taken on a greater importance in the past year. With employees from over 30 different countries, the finances and HR department is continually growing to meet the needs of each employee.
2. **Communication, Public Relations & Marketing (Nicolas Filippov)**: The mission of the communication, public relations and marketing department is to use all forms of media and communication to build, maintain, manage the reputation of the Institute, and to promote the Idiap services available for external institutions, such as EU project management, submission proposal tools, etc.
3. **System and Infrastructure (Frank Formaz)**:The main mission of the system and infrastructure group is to provide an optimal and efficient work environment for the Idiap collaborators. The tasks can be split into three main activities covering (1) centralized IT services for the whole Institute (network, storage, servers, workstations, high performance computing, identity management, data distribution), (2) support for collaborators (helpdesk, project specific tasks, web presence), and (3) Infrastructure (building, offices, equipments, central purchasing office).
4. **Technology Transfer (Dr. Joël Dumoulin)**: Technology transfer is one of the Idiap Research Institute's three core missions. One of the fundamental challenges is to facilitate the interface between the knowledge and the skills of the researcher and the needs of the industrial partner. Idiap resolves this by providing a dedicated multi-disciplinary team of developers and programmers which transfers pieces of software, algorithms, knowledge and expertise. This transfer of technology is usually done by granting rights on the commercial exploitation of this technology (through license).
5. **Program Managers (Justine Darioly & Barbara Huguenin)**: The work of the program management team is divided into two types of activities. The first is the provision of services to researchers within the framework of European and Swiss projects. The second category includes activities ranging from event organization to database management, which are not directly linked to the management of research projects but facilitate the work of Idiap researchers.
6. **Research and Development Engineers (Olivier Bornet)**: The mission of the research and development team is to provide support to Idiap researchers in the software development tasks. Their first mission is thus to second our research efforts by building prototypes, implementing and testing algorithms, designing and running experiments, and managing legacy code. Their second mission concerns all of our technology transfer tasks. The research and development group also provides daily support to Idiap researchers (software disclosures, showroom and internal demonstrators, development tools). Finally, the research and development team is involved in multiple key facets of our Master in Artificial Intelligence, including teaching, technical support to teachers, setup of new tools, like JupyterHub, and supervision of the students in their on-the-job activities.
7. **Legal Adviser (Marie-Constance Kaiflin)**: The main missions of the legal adviser are to write, analyse, negotiate project contracts (research, consortium, consultant agreements, NDA, Memorandum of Understanding, etc) or technology transfer contracts (patents, knowhow, licenses) with industries, universities or research institutions. The legal adviser deals also with all the legal aspects related to human resources (work contracts, staff regulations rules) and data protection (ethics, databases collection and distribution).

3 Research Groups

3.1 Speech and Audio Processing

Heads: Prof. Hervé Bourlard (MS and PhD, Polytechnic University, Mons, Belgium, 1982 and 1992), Dr. Philip N. Garner (MEng, University of Southampton, UK, 1991; PhD, University of East Anglia, UK, 2012), Dr. Mathew Magimai-Doss (MS by Research, Indian Institute of Technology Madras, India, 1999; PhD, Ecole Polytechnique Fédérale de Lausanne, Switzerland, 2005), Dr. Petr Motlicek (MS and PhD, Brno University of Technology, Czech Republic, 1999 and 2003).

Group overview: Speech processing has been one of the mainstays of Idiap's research portfolio for many years, covering most of the aspects of speech processing such as multilingual automatic speech recognition (ASR), speech synthesis, speech coding (including very low bit-rate), automatic speech intelligibility evaluation, or speech processing for analysis of motor speech disorders (e.g. pathological speech). The expertise and activities of the group encompass statistical automatic speech recognition (based on hidden Markov models (HMMs), or hybrid systems exploiting deep neural networks (DNN) and new end-to-end learning architectures), text-to-speech (TTS), speaker recognition (with extensions towards text-dependent and forensics scenarios) and generic audio processing (covering sound source localization, microphone arrays, speaker diarization, audio indexing, perceptual background noise analysis for telecommunication systems) and, more recently, compressive sensing, sparse recovery, and natural language understanding applied to ASR.

The Speech and Audio Processing group in 2022 was composed of 1 head of group, 3 principal investigators, 1 sabbatical academic visitor, 2 research associates, 9 postdocs, 12 PhD students, and 7 interns.

Key scientific outputs: Our primary research directions have traditionally been HMMs and DNN based approaches applied in acoustic modelling for various speech processing tasks. Techniques built around HMM and HMM-DNN architectures resulted in a unified approach used for automatic speech recognition, speech synthesis and other related classification tasks. The group is well placed to take full advantage of recent advances in new architectures of deep (i.e. end-to-end) learning (including large pre-trained acoustic or language models ready for subsequent fine-tuning to the application domain), studied in particular through *PyTorch, Kaldi* and other open source frameworks.

In 2022, several key research contributions were achieved by the group, including: (1) multilingual automatic speech recognition including new architectures deploying large pre-trained models, rapid cross-lingual adaptation, automatic speech recognition in low-resourced language conditions, with many applications toward language understanding, (2) new end-to-end architectures for speech recognition, intent detection, slot filling, partially targeting dialogue systems such as those deployed for contact centers, (3) speaker recognition, through both text-independent and particularly text-dependent (i.e. for speaker verification including forensic applications, or authorship attribution) scenarios and information fusion for large-scale speaker identification, (4) large scale media processing, including multilingual broadcast news recognition, and information retrieval from spoken documents, (5) new compressive sensing and Sparse Recovery theories to ASR, and dualities with sparse DNN auto-encoders, (6) detection of impairments in speech signal to uncover motor speech disorders, (7) paralinguistic speech processing with hand-crafted features and neural embeddings, (8) sign language recognition and assessment for sign language learning, (9) joint modeling of speech and physiological signals, (10) signal processing and machine learning for bioacoustic audio analysis, and (11) speech analytics specifically focusing on domain of air-traffic management, as well for security domain.

Beside that, the group is also involved in the deployment of speech (including subsequent tasks such as language understanding) and speaker recognition algorithms for industrial applications, and is regularly involved in international evaluation campaigns (such as those organised by National Institute of Standards

and Technology (NIST), U.S.).

Additional information and a list of projects are available from www.idiap.ch/speech.

Automatic speech recognition (ASR)

In recent years, our ASR research activities have been expanded from mono-lingual to cross-/multi-lingual processing. More specifically, in addition to focusing on "majority" languages other than English, French, or German, Idiap is actively carrying research in several ASR directions, including:

- **Robust parametrisation of ASR models**: We are investigating new features (e.g., posterior-based features) and new acoustic models (new forms of HMMs, or artificial neural networks, including very novel architectures such as Transformer or conformer based) that are more robust to noise and acoustic environments, as well as to speaker variability (e.g., accented speech, or dialect). In the context of the recently ended EC H2020 projects (ATCO2 and HAAWAII)[12], we were developing semi-supervised learning methods for rapid adaptation of speech recognition models to new (unseen) domains using unlabelled or partially labelled data.

- **Cross-lingual and multi-lingual speech recognition (including low-resource scenarios) with applications beyond speech**: Until 2022, Idiap has been collaborating on the US IARPA SARAL project[13]. As illustrated in Figure 2, the project aimed at developing cross-lingual spoken document retrieval and summarization techniques that will work for any language in the world, given minimal resources to work with. In those contexts, IDIAP focused on investigating and exploiting fast acoustic model adaptation techniques in cross-lingual and multi-lingual scenarios. Further, we collaborated with project partners on subsequent tasks processing automatically generated transcripts by NLP algorithms. The concept is also partially exploited in an industrial collaboration with Uniphore[14], a world leader in Conversational AI, to develop ASR engines and downstream language understanding technologies for specific non-European languages.

- **Swiss languages**: We continuously improve our multilingual speech recognisers for Swiss German and Swiss French and also apply the most recent advances in speech technology employing Deep Neural Networks (DNN) and end-to-end approaches. Since 2015, we collaborate with *recapp IT AG* on a wider range of Swiss dialects towards the first commercial product that performs Swiss German (dialect) speech recognition. From 2019 until late 2021, a collaboration with Swisscom enabled us to investigate "lexicon-free" advances in the field, which proved to be particularly suitable for dialects with no standardised orthography. This also led to work on code switching, where users change language (typically to English) during a sentence. Between 2019-2022, Idiap was leading activities on development of large-scale multilingual speech analytics engine (automatic speech recognition, language understanding including detection of entities, topics, summarisation, etc.) oriented toward law enforcement. The deployment with 10 European police end-users was done through EC H2020 ROXANNE project (see more in Section 3.1).

- **Exploiting compressive sensing and sparse recovering theories for ASR**: Through SNSF funded projects PHASER, PHASER-QUAD and SHISSM[15], Idiap is still researching new theoretical links between compressive sensing, sparse auto-encoders, and statistical/HMM-DNN approaches towards improving ASR performance and noise robustness.

[12] https://www.atco2.org, https://www.haawaii.de
[13] https://www.idiap.ch/en/scientific-research/projects/SARAL
[14] https://www.uniphore.com
[15] https://www.idiap.ch/en/scientific-research/projects/SHISSM

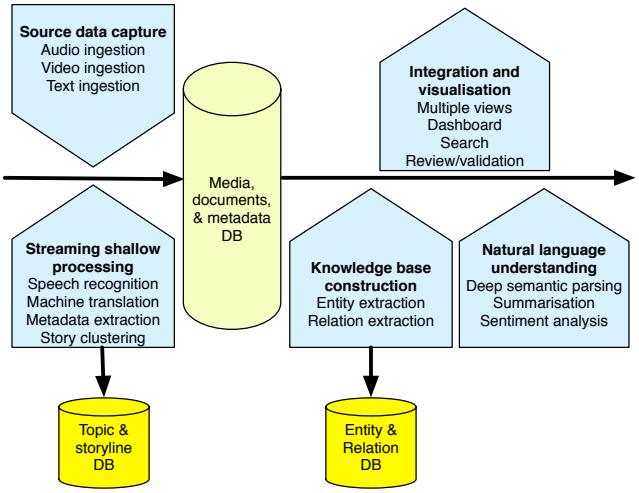


Figure 2: An overview of a typical multilingual multimedia processing stream, as used in ongoing US-DARPA SARAL project in the context of multilingual speech recognition at Idiap.

Text-to-speech synthesis (TTS)

TTS has been an established venture for the speech group at Idiap for perhaps fifteen years. TTS has been central to several past projects. The group has continued to track recent developments in deep learning, the most important being in generative approaches such as diffusion. Work on adaptation of diffusion was conducted under the NAST (neural architectures for speech technology) project. Research continued to focus on how to integrate emotional indicators based on prosody into state of the art deep learning solutions.

Speaker recognition and voice biometrics

Idiap is actively carrying R&D on significantly improving capabilities of voice technologies in biometric person identification applicable to very large scale data. Two activities were pursued in 2019-2022: (1) combining speaker recognition (including identification, diarization and clustering) with other tasks such as natural language understanding and network analysis to combat organized crime through ROXANNE H2020 project under Fighting Crime and Terrorism work programme[16], (2) improving the state-of-the-art speaker technologies by integrating recent advances in machine learning (especially through the participation on NIST speaker recognition evaluations[17]), and (3) working on specific biometrics technologies applied on audio such as authorship attribution from spoken material (e.g. allowing to identify authors of spoken documents).

[16] https://www.roxanne-euproject.org
[17] https://www.nist.gov/itl/iad/mig/speaker-recognition

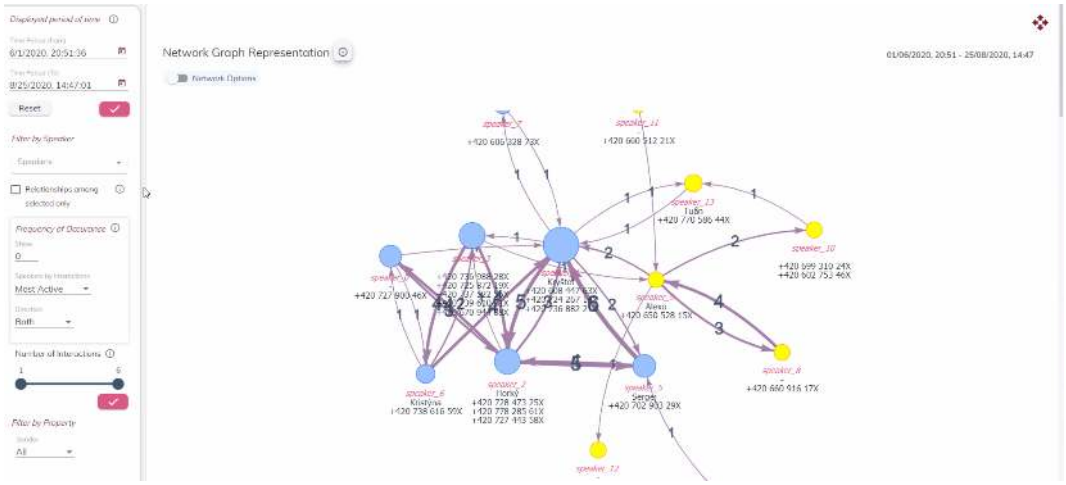


Figure 3: ROXANNE platform to process audio data and combining them with contextual information.

Pathological speech processing

Speech and language impairments can occur due to various reasons such as, due to neurological disorders, oral cancer, hearing loss. In recent years, Idiap has been actively involved in such impaired or pathological speech processing in collaboration with clinical researchers. The SNSF Sinergia project MoSpeeDi[18] (Motor Speech Disorder) focuses on (i) developing accurate models to characterize non-impaired and impaired phonetic speech planning and motor speech programming, and ii) developing automatic techniques to detect and classify several speech impairments as well as to assess the intelligibility of patients. The EU H2020 MSCA-ITN-ETN project TAPAS[19] which is targeting three key research problems, (1) detection, (2) therapy, and (3) assisted living so that it works well for people with speech impairments and also helps in making informed clinical choices. In the recently started SNSF Bridge Discovery project EMIL[20], we are contributing towards development of paralinguistic speech processing techniques for detection and monitoring of emotion and mood in Parkinson's disease patients, and integrating them with other ambient and wearable sensor modalities for development of a new closed-loop deep brain stimulation control system.

Other research directions

- **Physiological influences**: Under the NAST project and the new NCCR Evolving Language, we are investigating both how physiological processes of perception and production can influence speech technology, and how the quite mature technology can say something about our understanding of physiology. In 2022, two distinct threads continued: we investigate how to combine conventional DNNs with the spiking neurons thought to be closer to the physiological function, and we investigate the latent understanding of the ear and cochlea, including the "efferent" path. The former in particular has implications for driving future ultra-low power machine learning research.

- **Sign language processing**: In the framework of recently completed SNSF Sinergia project SMILE,[21] and its follow-up SNSF Sinergia project SMILE-II[22], Idiap is developing a sign language assessment system that can assist Swiss German sign language learners in standardizing a vocabulary production test to be aligned with levels A1 and A2 of the Common European Framework of Reference for Languages. Demonstration available at vimeo.com/297803984. Our current research activities

[18] https://www.idiap.ch/en/scientific-research/projects/MOSPEEDI
[19] https://www.tapas-etn-eu.org
[20] https://www.idiap.ch/en/scientific-research/projects/EMIL
[21] https://www.idiap.ch/scientific-research/projects/SMILE
[22] https://www.idiap.ch/en/scientific-research/projects/SMILE-II

are focusing on continuous sign language processing and extension of the SMILE technology to remote online sign language assessment.

- **Sound localization and microphone array**: Idiap continues to work on distant speech processing by contributing to the Perception and Activity Understanding group especially under industrial project activities. Recent work done through EC H2020 MuMMER project[23] focused on audio source localization, speech detection and speaker re-identification applied in robotics.
- **Higher level semantics**: Building on past work inferring charisma from text, we continued a project with the intention of inferring storytelling from audio recordings of selection interviews. The SNSF-funded SteADI project is a collaboration with the universities of Lausanne and Neuchâtel.
- **Joint acquisition and modeling of speech and physiological signals**: In recently funded SNSF project TIPS[24], Idiap is collaborating with CSEM[25] to develop a platform where speech and physiological signals are collected in a synchronous manner through a wearable cooperative sensor and processed to develop novel speech- and physiology-based applications. Our current research in this direction has been focusing on speech-based breathing pattern estimation using neural networks and exploitation of such pre-trained neural networks for auxiliary speech tasks such as, speech-based Covid detection.
- **Paralinguistic speech processing**: In the context of Innosuisse project CMM, SNSF Discovery project EMIL and HASLER Foundation project FLOSS, Idiap has been working on various paralinguistic speech processing tasks, namely, prediction of non-expert perceived speech fluency ratings, perceived speaker personality prediction, turn level emotion prediction and speech pseduonymization.
- **Text processing (language understanding) algorithms applied on automatically generated transcripts**: Many industrial applications require not only to automatically recognize the input speech with a high accuracy but also apply (or ideally combine with) subsequent modules (such as summarisation, entity recognition, retrieval, intent detection, polarity detection) on the output of speech multilingual recognition. Among typical tasks we can count word boosting enabling to significantly increase recognition accuracy of apriori pre-defined words or entities (known to the user in advance).
- **Applications to the air-traffic management and communication intelligence**: Several European and industrial projects have addressed in 2022 the domain of air-traffic communication as well the domain of signal and speech analytics for the communication intelligence (with Armasuisse[26]).
- **Detecting Cognitive Processes and Mental Health Conditions**: In 2022, several works were addressing a challenge in domain of psycholinguistics, motivated by revealing information about the use of language to uncover personality of an individual targeted by social and medical sciences.

Key publications

[1] A. Bittar and P. N. Garner, “A surrogate gradient spiking baseline for speech command recognition,” *Frontiers in Neuroscience*, Aug. 2022

[2] S. P. Dubagunta, R. J. J. H. van Son, and M. Magimai.-Doss, “Adjustable deterministic pseudonymization of speech,” *Computer, Speech, and Language*, vol. 72, Mar. 2022, Open Access

[3] E. VILLATORO-TELLO, S. Madikeri, P. Motlicek, *et al.*, “Expanded lattice embeddings for spoken document retrieval on informal meetings,” in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Jul. 2022

[23] http://mummer-project.eu
[24] https://www.idiap.ch/en/scientific-research/projects/TIPS
[25] www.csem.ch
[26] https://www.ar.admin.ch/en/home.html

[4] S. Burdisso, Z.-G. Juan, E. VILLATORO-TELLO, *et al.*, “Idiapers @ causal news corpus 2022: Efficient causal relation identification through a prompt-based few-shot approach,” in *The 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE @ EMNLP 2022)*, ACL Anthology (https://aclanthology.org/2022.case-1.9), Dec. 2022

3.2 Social Computing

Head: Prof. Daniel Gatica-Perez (PhD, University of Washington, USA, 2001; EPFL Adjunct Professor)

Group overview:

The social computing group studies how people and technology interface in everyday life. The group's research combines theories and methods from computing (ubiquitous computing, social media, machine learning) and the social sciences to analyze human and social behavior and design systems that support individuals and communities.

The Social Computing group in 2022 involved the group head, four PhD students, two part-time developers, one visiting PhD student, and eleven student interns. The main research lines investigated in 2022 included mobile crowdsensing for health and cities; human-centered analysis of social media and news; and privacy-preserving machine learning.

Key scientific outputs: Publications on diversity-aware mobile crowdsensing to characterize everyday life activities; human-centered news and social media analysis; and privacy-preserving machine learning. 16 EPFL PhD students have graduated from the group since 2002.

Additional information and a list of projects are available from www.idiap.ch/socialcomputing.

Mobile crowd-sensing for health and cities

In the context of the European H2020 WeNet project[27], we are developing diversity-aware algorithms for mobile sensing to support the well-being of young adults, through a series of large-scale experiments in several European universities, as well as universities in Latin America and Asia. One key motivation of this work is the advocacy for diversity in data and algorithms to improve the representation of non-western citizens. More specifically, we conducted comparative studies across eight countries, focused on the automatic recognition of health-related variables from smartphone sensors. This includes variables like everyday life activities, self-reported mood, and eating habits (Figure 4). Our research deals with fundamental issues like model generalization across countries [1,2]. This research was complemented with the development of privacy-preserving machine learning methods that integrate differential privacy with graph neural networks, and that could potentially be used as part of future mobile and social applications [3].

Regarding mobile technologies for social innovation, in the European H2020 ICARUS project[28], we are investigating human-centered approaches that can support European cities to address some of their urban security priorities. This work requires a combination of social and technological innovation approaches, including design thinking, participatory research, and human-centered adaptation of computational tools. This resonates with our long-standing experience using the Civique platform, which allows to collect mobile data for local causes[29]. Civique has been used in a variety of applications, ranging from supporting cities to collect information related to urban issues like street harassment to teaching students about humanitarian technologies, and more recently to capture the experience of Swiss residents during the COVID-19 lockdown as part of a multi-disciplinary citizen science effort [4].

Human-centered analysis of social media and news

In the context of the European H2020 AI4Media project [30], we are investigating human-centered approaches to understand European news consumption across multiple information channels, with focus

[27] https://www.internetofus.eu
[28] https://www.icarus-innovation.eu
[29] https://www.civique.org
[30] https://www.ai4media.eu

on local news sources and health-related topics. More specifically, among other research lines we conducted a study over multiple European countries and newspapers to understand how the important issue of COVID-19 vaccination was treated over a period of two years using NLP approaches [5]. Furthermore, we published research on how the analysis of health-related videos shared in platforms like YouTube can complement the work of health psychologists studying new practices of health promotion (in collaboration with the University of Lausanne) [6].

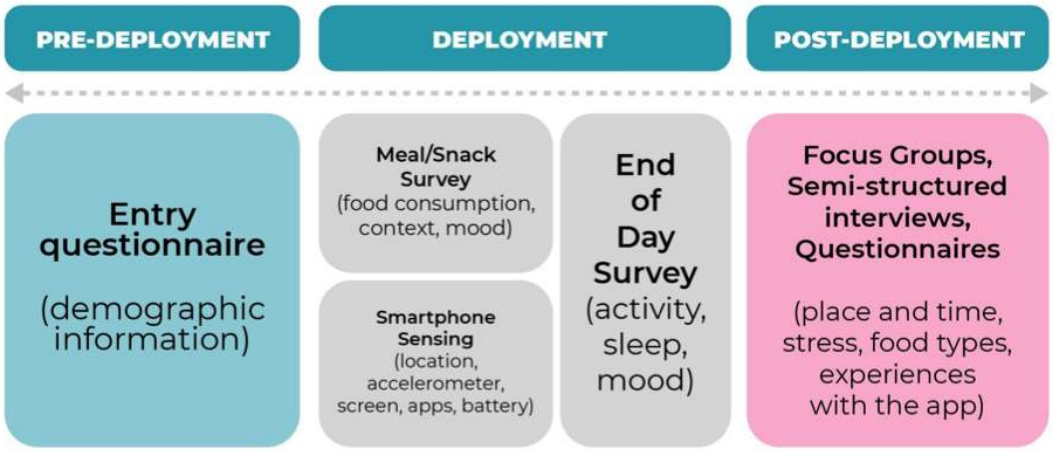


Figure 4: Research example: smartphone-based study design to analyze food consumption.

Key publications

[1] L. Meegahapola, W. Droz, P. Kun, A. de Gotzen, C. Nutakki, S. Diwakar, S. Ruiz Correa, D. Song, H. Xu, M. Bidoglia, G. Gaskell, A. Chagnaa, A. Ganbold, T. Zundui, C. Caprini, D. Miorandi, A. Hume, J. L. Zarza, L. Cernuzzi, I. Bison, M. D. Rodas Britez, M. Busso, R. Chenu-Abente, F. Giunchiglia, C. Gunel, L. Schelenz, and D. Gatica-Perez, Generalization and Personalization of Passive Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries, *PACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies (IMWUT)*, Vol. 6, No. 4, Dec. 2022.

[2] W. Bangamuarachchi, A. Chamantha, L. Meegahapola, S. Ruiz-Correa, I. Perera, and D. Gatica-Perez, Sensing Eating Events in Context: A Smartphone-Only Approach, *IEEE Access*, Vol. 10, pp. 61249-61264, Jun. 2022.

[3] S. Sajadmanesh, A. S. Shamsabadi, A. Bellet, and D. Gatica-Perez, GAP: Differentially Private Graph Neural Networks with Aggregation Perturbation, in *Proc. USENIX Security Symposium,* accepted for publication.

[4] L. Fritz, U. Vilsmaier, G. Clement, L. Daffe, A. Pagani, M. Pang, D. Gatica-Perez, V. Kaufmann, M. Santiago Delefosse and C. R. Binder, Explore, Engage, Empower: Methodological Insights into a Transformative Mixed Methods Study Tackling the Covid-19 Lockdown, *Humanities and Social Sciences Communications*, Springer-Nature, Vol. 9, No. 175, May 2022.

[5] D. Alonso del Barrio and D. Gatica-Perez, How Did Europe's Press Cover Covid-19 Vaccination News? A Five-Country Analysis, in *Proc. ACM International Workshop on Multimedia AI against Disinformation*, Newark, Jun. 2022.

[6] T.-T. Phan, C. Michoud, L. Volpato, M. del Rio Carral and D. Gatica-Perez, Health Talk: Understanding Practices of Popular Professional YouTubers, in *Proc. Int. Conf. on Mobile and Ubiquitous Multimedia (MUM)*, Lisbon, Nov. 2022.

3.3 Machine Learning

Head: Dr. Damien Teney (M.Sc. University of Liège, 2009; PhD, University of Liège, 2013)

Group overview: Machine learning encompasses computer techniques that modulate their behavior according to exemplar data. It has resulted in technologies at the core of many modern every-day data-processing software and apparatus. The objective of the Machine Learning group is to develop novel machine-learning techniques of general use, applicable in particular to complex visual and textual data. The group focuses in a particular on methods able to cope with artefacts of real-world data such as noisy annotations, distribution shifts, spurious correlations, and generalization across environments. This research can be motivated by general, fundamental problems, as well as concrete industrial applications or use cases.

The machine learning group at Idiap was formerly headed by Prof. François Fleuret, now at the University of Geneva. Dr. Damien Teney moved to Idiap in May 2021 and is the current head of the group. The group had been composed on average during the recent years of four PhD students, and one or two developers working on industrial applications. Current research in the group addresses a new range of topics in fundamental and applied machine learning outlined below.

Key scientific outputs: In 2022, the machine learning group presented multiple publications at top-tier conferences in machine learning and computer vision. This work contributed to improving the fundamental understanding of deep learning techniques [4], [8]. We also developed practical solutions to specific tasks such as image retrieval, and question-answering systems capable of reasoning over text and images simultaneously [6], [7]. We also developed a state-of-the-art technique for active learning [5], a paradigm of great practical interest since it can reduce the costs of acquiring and labeling training data in industrial applications.

Additional information and a list of projects are available from www.idiap.ch/ml.

Robustness and generalization.

The field of machine learning has made enormous advances over the past decade. Techniques based on deep learning in particular have enabled a number of applications, but there remains however a lack of understanding of the mathematical foundations of these techniques. Our work aims to improve the understanding of the limits of applicability of deep learning, especially regarding generalization across datasets, environments, and distribution shifts. For example, data collected in varying conditions, such as images captured under different lighting conditions or with different cameras, can adversely affect the effectiveness of image recognition models. The quest for models resilient to such variations is both a practical and a fundamental one. Our recent work helped understand why deep learning techniques are prone to capture spurious correlations in training data, and how to better guide the learning process to focus on patterns that better generalize at test time.

Vision & language as a test-bed for machine learning.

We use various tasks at the intersection of computer vision and natural language processing as test-beds to evaluate our new techniques. Vision-and-language tasks constitute one of the most active and challenging application areas of machine learning. For example, we have developed models capable of automatic image captioning and visual question answering (VQA) (see Figure 5). These tasks are challenging because they require extracting and relating semantic contents from both text and images. These contributions are touching fundamental research topics in machine learning, such as multimodal embeddings, compositional generalization, while delivering techniques directly useful for the real world. Our contributions on visual question answering, for example, have been applied to medical images to help experts navigate through large repositories of medical imagery via text queries in natural language.

Figure 5: We use various tasks at the intersection of computer vision and natural language processing as a test-bed for our machine learning methods. We have developed models capable of automatic image captioning (left) as well as answering general questions about user-provided images (right). These tasks are extremely challenging since they require to extract and relate semantic contents from different modalities. Our models are entirely learned from data (e.g. hundreds of thousands of questions with their correct answer) rather than human-designed reasoning mechanisms.

Beyond statistical learning.

Most of mainstream machine learning is essentially based on learning statistical patterns, which has inherent limitations that become increasingly apparent as the tasks addressed become more and more complex. Our research investigates complementary approaches that could allow to break through these limitations. Causal reasoning offers a framework to extract models of the world beyond mere correlations. They allow drawing inferences about the effect of interventions and about the mechanisms behind the observed data. Meta learning is another line of research that attempts to "learn to learn", i.e. optimize the learning process itself. These techniques are very active and promising research areas that our group is contributing to.

Key publications

[1] D. Teney, E. Abbasnejad, S. Lucey, *et al.*, "Evading the simplicity bias: Training a diverse set of models discovers solutions with superior ood generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022

[2] I. Hwang, S. Lee, Y. Kwak, *et al.*, "Selecmix: Debiased learning by contradicting-pair sampling," in *Advances in Neural Information Processing Systems*, 2022

[3] A. Parvaneh, E. Abbasnejad, D. Teney, *et al.*, "Active learning by feature mixing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022

[4] V. Shevchenko, D. Teney, A. Dick, *et al.*, "Reasoning over vision and language: Exploring the benefits of supplemental knowledge," in *ArXiv*, 2022 (under review)

[5] V. Shevchenko, E. Abbasnejad, A. Dick, *et al.*, "Ebms vs. cl: Exploring self-supervised visual pretraining for visual question answering," in *ArXiv*, 2022 (under review)

3.4 Perception and Activity Understanding

Head: Dr. Jean-Marc Odobez (PhD, INRIA/Rennes University, France, 1994; EPFL MER)

Group overview: The group investigates novel computational learning methods from computer vision, multimodal signal processing, and machine learning for the analysis of human activity from multi-modal data. Human and human-robot interaction analysis, surveillance, traffic monitoring or multimedia content analysis are the main application domains.

Over the last 3 years, the group was composed on average of one associate researcher, two post-doctoral researchers and three PhD students, one research engineer from the development team.

Key scientific outputs: The group is known for its work on multi-object tracking, temporal motif discovery and non-verbal behavior extraction. In particular, its patended work on 3D face and gaze tracking has led to the creation of the Eyeware SA company in 2016, and the team ranked second at the ICCV 2019 Facebook Synthetic Eye Generation Challenge. In recent years, the group has investigated deep learning methods for several tasks like gesture recognition, audio-visual speaking activity modeling, gaze, audio localization and speech/non-speech detection, body landmark detection, and multimedia processing (cross-modal transfer learning, shape recognition, text localization and semantic categorization). The group also integrates its algorithms into real-time perceptual systems used in collaborative projects or by companies: an HRI system for the Pepper robotic platform in the EU MuMMER project, see video; an anti-tailgating detection system (for Fastcom); a press prediction system (innoswiss project P3 finished in 2022, with Constellium), helping engineers to more quickly assess the manufacturability of an aluminium shape profile requested by a customer; an emotion recognition method for job interviews (innoswiss ADVANCE project, featured in the national TV[31]). During the period 2018-2022, the group published 12 journal papers and 28 conference papers, and filled 2 patents.

Additional information and a list of projects are available from www.idiap.ch/perception.

Human activity analysis, non-nerbal behavior (NVB) extraction, human-robot interactions

Head and body pose inference. The team has developed a robust and accurate head pose tracking framework from RGB-D data combining the benefits of the online fitting of a 3D face morphable model with the online 3D reconstruction of the full head Research on 2D and 3D body pose estimation is also conducted like investigating the use of transformer models for pose predictions [1].

Gaze analytics. Due to visually unobservable gaze variabilities across people, we have investigated several methods for building user-specific 3D gaze direction prediction models from user samples. Recently, we followed the idea of using pairs of eye images of the same person as input to predict gaze differences (differential gaze), and showed superior performance over the state-of-the-art [2]. We also demonstrated the possibity of designing a fully unsupervised gaze representation learning approach, allowing to leverage internet-scale data, by jointly learning a gaze representation network and gaze redirection network (see Fig. 6c) only from pairs of eye images, using as task the redirection synthesis of the second image from the first image. In the context of the ROSALIS SNSF project, we showed that gaze coordination priors involved in object manipulations and speaker interaction could be leveraged for obtaining weakly labeled gaze samples and robustly estimate gaze calibration models [3].

Gaze prediction in the wild. We are also investigating the gaze following task, a generalization of the gaze direction prediction task. It aims at predicting, in arbitrary scene images, where (defined as the image 2D location) a person in the image is looking at. We proposed different methods to address this task, exploiting for instance prior cues like pose and depth (see Fig. 6) [4]. As our research will be applied within the context of the AI4Autism SNSF project, which is about studying the digital phenotyping and profiling of autism spectrum disorders in children (attention deficit is a strong marker of ASD), we have also collected a large dataset (ChildPlay) of children data annotated with gaze information to evaluate our system.

[31] Mise au point, TSR, 18/10/2020.

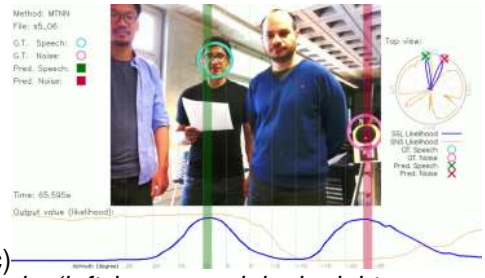
a)  b)  c) 

Figure 6: Human activity analysis. a) Redirected gaze example (left image: original; right one: redirected gaze towards the nose). b) Gaze following. 3D gaze direction prediction (arrow and blue circle) for the child, with possible targets (gold highlight) and 2D gaze prediction (green point). c) Multiple sound source localization and discrimination with a multi-task network predicting for each direction (circle on the right) the sound source likelihood. Sources are then mapped to visual data.

Multimodal analysis, robotics, tech transfer

Audio analysis. We studied different DNN architectures and unsupervised methods for sound processing: efficient multi-task approach for the joint localization and categorization (speech vs non-speech) of multiple sources from a microphone array (Fig. 6), along with speaker embeddings.

Manipulation task policy learning. In the context of the EU HEAP project, we have investigated novel learning methods for Deep Q-learning. In particular, we studied the case of pushing tasks (Fig. 7a), an ubiquitous manipulation task in object handling, and demonstrated that the use of proper image-to-image translation architecture, better exploitation of the reward structure during training, as well methods for encoding task parameters, were improving current techniques.

Agriculture. We have extended our know-how on landmarks prediction to the challenging case of predicting the structure of vine trees (Fig. 7). The output will be exploited by our 3D2Cut partner within augmented reality glasses to assist operators in some decision-making activities like grapevine pruning.

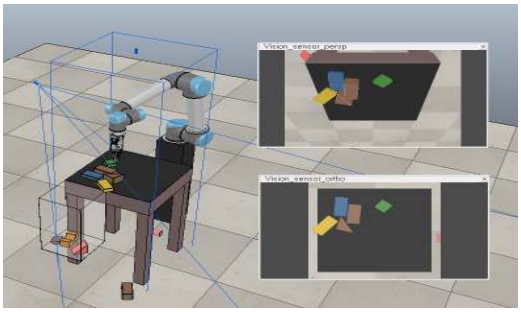
a)  b) 

Figure 7: a) Investigating novel robotics visual processing architectures (here, pushing into a box). b) Vine structure prediction: localization of nodes, branches, and prediction of their type. Left image: prediction; right: ground-truth.

Key publications

[1] A. Martínez, M. Villamizar, and J.-M. Odobez, "Pose Transformers (POTR): Human Motion Prediction With Non-Autoregressive Transformers," in *Int. Conf. on Comp. Vision (ICCV) Workshops*, 2021.

[2] G. Liu, Y. Yu, K. Funes and J.-M Odobez. "A Differential Approach for Gaze Estimation," in *IEEE Trans. Pattern Anal. Machine Intelligence (PAMI)* , Vol 43(3), 2021.

[3] R. Siegfried and J.-M. Odobez. "Robust Unsupervised Gaze Calibration using Conversation and Manipulation Attention Priors," *ACM Trans. on Multimedia Computing, Communications, and Applications*, Vol 18(1), pp 20:1-20:27, 2022.

[4] A. Gupta, S. Tafasca and J.-M. Odobez. "A Modular Multimodal Architecture for Gaze Target Prediction: Application to Privacy-Sensitive Settings," *IEEE CVPR Workshops*, 2022.

[5] T. Gentilhomme, M. Villamizar, J. Corre and J.-M. Odobez. "Towards smart pruning: ViNet, a deep-learning approach for grapevine structure estimation," *JI of Computers and Electronics in Agriculture*, 2023.

3.5 Genomics and Health Informatics

Head: Dr. Raphaëlle Luisier (Master of Science in Bioengineering and Biotechnology, EPFL, 2009 & PhD in Bioinformatics from the Basel University, 2013)

Group overview: The Genomics & Health Informatics Group was created in 2019 with the overarching goal to transform our understanding and treatment of devastating and incurable human diseases by adapting cutting-edge computational methods inspired from various signal processing fields such as computer vision and language processing to longitudinal biological and clinical data. The group works in close collaboration with biologists and clinicians from hospitals (CHUV, Sion Hospital), academia (Francis Crick Institute, London; Kings College, London; University College London) and industry (NIBR, Suva), thereby maintaining a fertile ground for innovation, learning and discoveries with real therapeutic prospects. Since December 2020, the group is an associate member of the NCCR RNA & Disease. Dr Luisier is also a group leader at the Swiss Institute of Bioinformatics (SIB) since June 2022. The group has expertise in genomics, bioinformatics, RNA biology, neuroscience, data science, and data visualisation.

The group is also taking part to frequent outreach activities in order to engage the public on the scientific activities led at Idiap and also encourage young women to pursue a career in science. Such activities include, but are not limited to, interviews for radio channels (Radio Chablais, RTS CQDF), mentoring young ladies within the "SwissTech ladies" (SATW) framework, and interview for "TTC Le valais Innovant".

Key scientific outputs: Since its establishment in December 2019, the Genomics & Health Informatics group has developed computational and machine learning methods to histopathological images and cellular imaging that enable the automated extraction of key information from such complex images leading to key discoveries related to neurodegenerative disorders. The methods and associated discoveries were published in leading peered-review journals in neuroscience and pathology, as briefly described below.

Additional information and a list of projects are available from www.idiap.ch/genomics.

Deep learning method to robustly and efficiently test biological hypotheses

By deploying transfer learning with a pre-trained deep-learning image classifier to a rich multichannel fluorescence microscopy high content imaging data-set, the group revealed novel neurite-associated morphological perturbations in ALS disease. This method, published in *Neuropathology and Applied Neurobiology* (Verzat et al. 2021) allows to robustly and efficiently test biological hypotheses in an unbiased fashion and therefore is potentially transformational in terms of the novel phenotypic information that can be gleaned from image analysis of in vitro neurodegeneration models. In line with Idiap's incentive to promote reproducibility in research, the method has been publicly released on Gitlab (ALS classification) and the large-scale imaging dataset of 156,577 images has been publicly released in the Image Data Resource (IDR) under study (idr0112).

The Genomics & Health Informatics group in 2022 was composed of 2 EPFL MSc students, 1 EPFL BSc student, 3 interns, and 2 PhD students.

Advance in histopathological tissue sections analysis

Histopathological analysis of tissue sections is an invaluable resource in neurodegeneration research. However, cell-to-cell variation in both the presence and severity of a given phenotype is a key limitation of this approach. The Genomics & Health Informatics group directly addressed these issues by combining automated image processing with machine learning methods to substantially improve the speed and reliability of identifying phenotypically diverse *Motor Neurons* (MNs) populations. The developed method enabled automatic identification of sick cells with unprecedented accuracy and at single cell resolution in histopathological tissue sections, and was published in *Brain Pathology* (Hagemann et al., 2021). This study showcase the potential of combining histopathology with automated image processing and machine

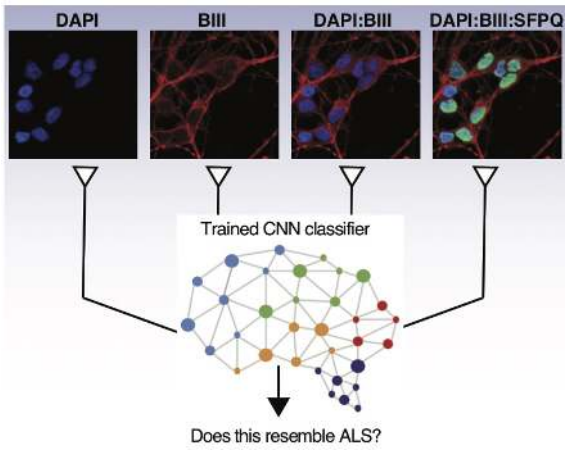


Figure 8: Consecutive training of deep learning image classifiers using a combination of several fluorescent biomarkers to discriminate control from ALS motor neuron cultures enables to identify disease cellular compartment. (fig. from C.V. et al. *Neuropathology and Applied Neurobiology, 2021)*

learning and might prove transformational in our understanding of ALS and neurodegenerative diseases more broadly.

Discovery of a novel hallmark of Amyotrophic Lateral Sclerosis disease

Amyotrophic lateral sclerosis (ALS) is a rapidly progressive and incurable neurodegenerative disease, characterized by the progressive death of motor neurons (MNs) for which the early cellular and molecular events remain poorly understood. As a consequence, no effective treatments have been developed for this devastating disease where most patients die within 3-5 years (with 30% dying in just 12 months of diagnosis). The Genomics & Health Informatics group is working in close collaboration with the laboratory of Rickie Patani (The Francis Crick Institute, London) in order to identify the cellular and molecular events underlying the disease. In the last two years, the group has developed bioinformatic pipelines to RNA-sequencing data generated in Patani's laboratory. These methods enabled to demonstrate a progressive accumulation of intronic sequences in the cytoplasm of developing motor neurons *Brain* (Tyzack et al, 2021), together with a potential mechanism in the disease pathogenesis with real therapeutic potential (see Petric-Howe et al, 2022 in *Genome Research*).

Key publications

[1] M Petric-Howe, H Crerar, J Neeves, J Harley, GE Tyzack, P Klein, A Ramos, R Patani* and R Luisier*. "Physiological intron retaining transcripts in the cytoplasm abound during human motor neurogenesis," in *Genome Research* (2022). *These authors contributed equally.

[2] C Verzat, J Harley, R Patani and R Luisier. "Image-based deep learning reveals the responses of human motor neurons to stress and VCP-related ALS," in *Neuropathology and Applied Neurobiology* (2021).

[3] C Andreassi*, R Luisier*, H Crerar, S Franke, NM Luscombe, G Cuda, M Gaspari and A Riccio. "3' UTR Remodelling of Localized Transcripts in Sympathetic Neurons Axons," in *Cell Reports* (2021). *These authors contributed equally.

[4] GE Tyzack, J Neeves, H Crerar, P Klein, O Ziff, DM Taha, R Luisier*, NM Luscombe* and R Patani*. "Aberrant cytoplasmic intron retention is a blueprint for RNA binding protein mislocalization in VCP-related amyotrophic lateral sclerosis," in *Brain* (2021). *These authors contributed equally.

[5] C Hagemann, GE Tyzack, DM Taha, H Devine, L Greensmith, J Newcombe, R Patani*, A Serio* and R Luisier*. "Automated and unbiased classification of motor neuron phenotypes with single cell resolution in ALS tissue," in *Brain Pathology* (2021). *These authors contributed equally.

3.6 Robot Learning and Interaction

Head: Dr. Sylvain Calinon (MS and PhD, EPFL, 2003 and 2007)

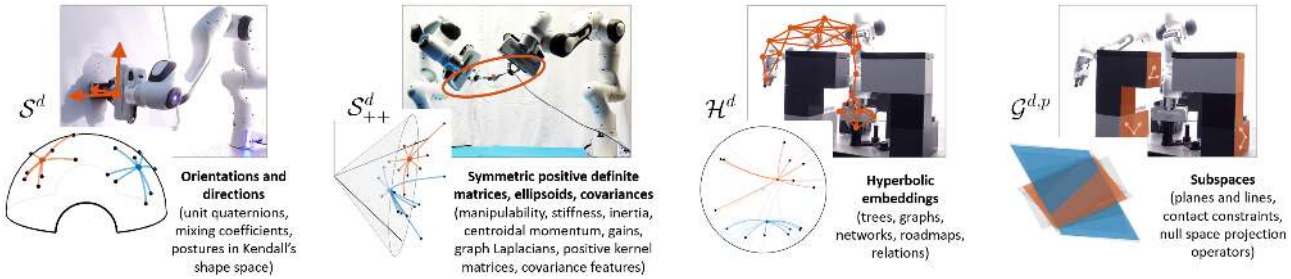
Group overview: The Robot Learning and Interaction group, created in 2014, focuses on human-centered robotics applications in which the robots can acquire new skills from only few demonstrations and interactions. It requires the development of models that can exploit data structures in an efficient way, the development of optimal control techniques that can exploit the learned task variations and coordination patterns, and the development of intuitive interfaces to acquire meaningful demonstrations. The developed approaches can be applied to a wide range of manipulation skills, with robots that are either close to us (assistive and industrial robots), parts of us (prosthetics and exoskeletons), or far away from us (teleoperation).

The Robot Learning & Interaction group in 2022 was composed of 3 postdoctoral fellows, 10 PhD students and 4 visiting PhD students.

Key scientific outputs: Development of learning techniques that only need a small number of demonstrations, by exploiting structures that can be found in a wide range of robotic tasks, and by exploiting bidirectional human-robot interaction as a way to collect better data.

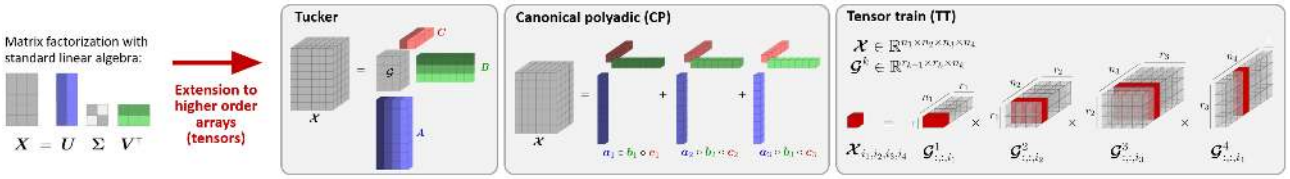
Additional information and a list of projects are available from www.idiap.ch/rli.

Geometry-aware learning and control



Data in robotics are characterized by simple but varied geometries, which are often underexploited when developing learning and control algorithms. We exploit Geometric Algebra and Riemannian geometry to extend algorithms initially developed for standard Euclidean data, by taking into account the structures of these manifolds to learn from small sets of data.

Tensor factorization in robotics applications



Another structure that we exploit relates to the organization of data as multidimensional arrays (also called tensors). Such data appear in robotic tasks either as the natural organization of sensory/motor data (tactile arrays, images, kinematic chains), or as the result of standardized preprocessing steps (moving time windows, covariance descriptors). Tensor methods extend standard factorization techniques such as singular value decomposition to multilinear decompositions, without requiring the transformation of the tensors into matrices or vectors.

This factorization approach can be applied to learning and optimization problems characterized by two

types of variables: 1) *task parameters* representing the situation that the robot encounters (locations of objects, users or obstacles); and 2) *decision variables* related to actions that the robot takes (controller acting within a given time window). For each change of task parameters, decision variables need to be recomputed as fast as possible, so that the robot can fluently collaborate with users and can swiftly react to changes in its environment. We investigate the roles of offline and online learning optimization to attain such objective, which is formalized as an *optimal control problem* with a cost function to minimize, parameterized by task parameters and decision variables. We investigate the use of *tensor train (TT) decomposition* as a model to learn the structure between the task parameters, the decision variables and the resulting cost expressed in the form of a probability distribution, which then allows solutions to be sampled from a conditional distribution. We exploit this structure to gather prior knowledge in an offline phase, which is further used for fast online decision making.

Combination of controllers as a product of experts

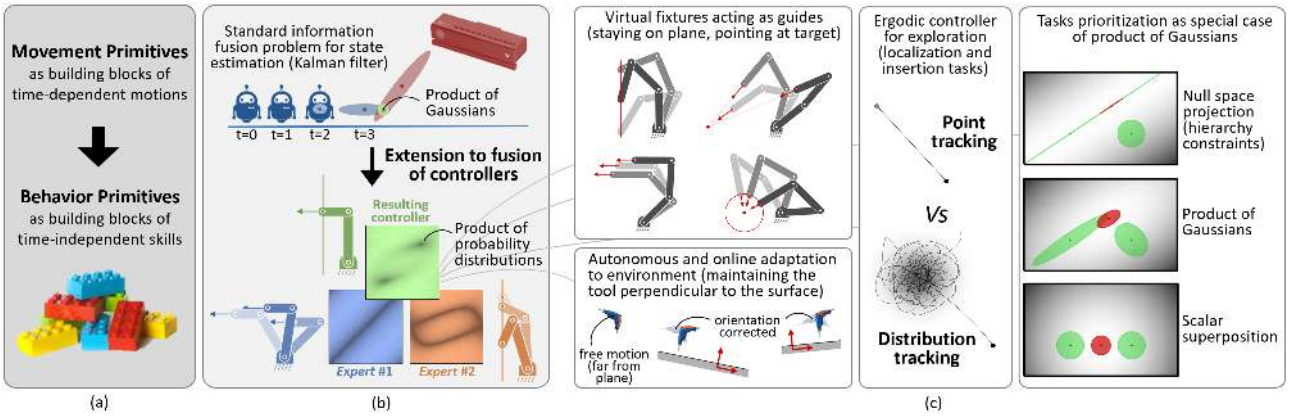


Figure 9: (a) *We extend the principle of movement primitives (superposition of basis functions as time-dependent trajectories) to behavior primitives (superposition of time-independent controllers).* (b) *State estimation in robotics is typically formulated as an information fusion problem. We propose to combine controllers in the same principled way by using a product of probability density functions, resulting in a shared control formulation in which as set of controllers assist the user to achieve the task.* (c) *We develop behavior primitives in both position and force domains, including the autonomous adaptation to objects, exploration behaviors, and tasks prioritization.*

We formulate the problem of combining controllers as a product of experts (see Fig. 9), where each expert takes care of a specific aspect of the task to achieve, which can be learned separately or altogether (by variational inference). With this probabilistic formulation, the robot counteracts perturbations that have an impact on the task, while ignoring other perturbations. This approach creates bridges with research in biomechanics and motor control, including minimal intervention principles, uncontrolled manifolds or optimal feedback control.

Key publications

[1] Shetty, S., Silvério, J. and Calinon, S. (2022). Ergodic Exploration using Tensor Train: Applications in Insertion Tasks. IEEE Trans. on Robotics (T-RO), 38:2, 906-921.

[2] Pignat, E., Silvério, J. and Calinon, S. (2022). Learning from Demonstration using Products of Experts: Applications to Manipulation and Task Prioritization. International Journal of Robotics Research (IJRR), 41:2, 163-188.

[3] Ti, B., Gao, Y., Zhao, J. and Calinon, S. (2022). Imitation of Manipulation Skills Using Multiple Geometries. In Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS), pp. 7391-7398.

3.7 Natural Language Understanding

Head: Dr. James Henderson (BSc, Massachusetts Inst. Technology, USA, 1987; MSE & PhD, Univ. Pennsylvania, USA, 1991,1994; MER & Chargé de Cours, Univ. Geneva, 2008–2012,2012–2018)

Group overview: The Natural Language Understanding group (NLU) works at the intersection of machine learning and natural language processing, with an emphasis on representation learning for the meaning of language, attention-based deep learning models, and structured prediction. We model summarisation, machine translation, knowledge extraction, syntactic and semantic structure, and semantics, among other natural language processing problems (NLP). We develop deep learning models of the discovery and prediction of entities and their relations at multiple levels of representation for multiple tasks.

During 2022, the NLU group had the following members: the head of the group, and 7 PhD students. Four of these students spent some of the year doing internships.

Key scientific outputs: During 2022, the work of the NLU group has produced several key publications. On a mature line of work, we published our work on using our Graph-to-Graph Transformer architecture for coreference resolution. On the topic of summarisation, we received the best resource paper award at the GEM workshop, where we also published our work on hallucination detection. On the topic of representation learning, we published one paper on learning mappings in the bag-of-vector latent space of Transformers, and released two arxiv papers, one on replacing the attention mechanism of Transformers for greater efficiency, and one on a variational Bayesian formalisation of the latent space of Transformers. We expect both these recent ideas to have a big impact in 2023.

Additional information and a list of projects are available from www.idiap.ch/nlu.

Deep Learning Architectures for Graphs

Deep learning models based on self-attention, in particular Transformer, have revolutionised the state-of-the-art in many NLP tasks, particularly when combined with pretraining (e.g. BERT). We have developed a version of Transformer which maps graphs to graphs, instead of sequences to sequences. Given the nodes of a graph, such as the words of a sentence, Graph2Graph Transformer can input arbitrary graphs and output arbitrary graphs over these nodes, using the self-attention mechanism, including the ability to iteratively refine the predicted graph (Figure 10(a)). We have previously shown state-of-the-art results for syntactic parsing, a core benchmark for structured prediction and NLU, and in 2022 we published our state-of-the-art results in coreference resolution.

Efficiency in Deep Learning Architectures

The scale of current NLP models, datasets, and even individual texts make efficiency a major issue in NLP. The above work on coreference resolution needed to propose solutions to fitting long documents in standard processing frameworks. Our work on efficient few-shot learning with language models proposed solutions to efficiently fine-tune very large pretrained language models. Our recent work on the "Hypermixer" architecture replaces the quadratic-complexity attention mechanism of Tansformers with a linear-complexity solution using MLPs generated by a shared hypernetwork.

Representation Learning for NLP Tasks

Deep learning is fundamentally about how to induce hidden representations. In recent years this has shifted from learning vector-space representations to learning attention-based (bag-of-vector space) representations. Our group has been at the forefront of understanding this shift, including the above Graph-to-Graph Transformer and Hypermixer architectures. In 2022, we extended our previous work on learning sequence-to-sequence tasks in the latent vector-space of a standard auto-encoder to a latent bag-of-vectors space, so that this method can be used with a Transformer auto-encoder. We also continued

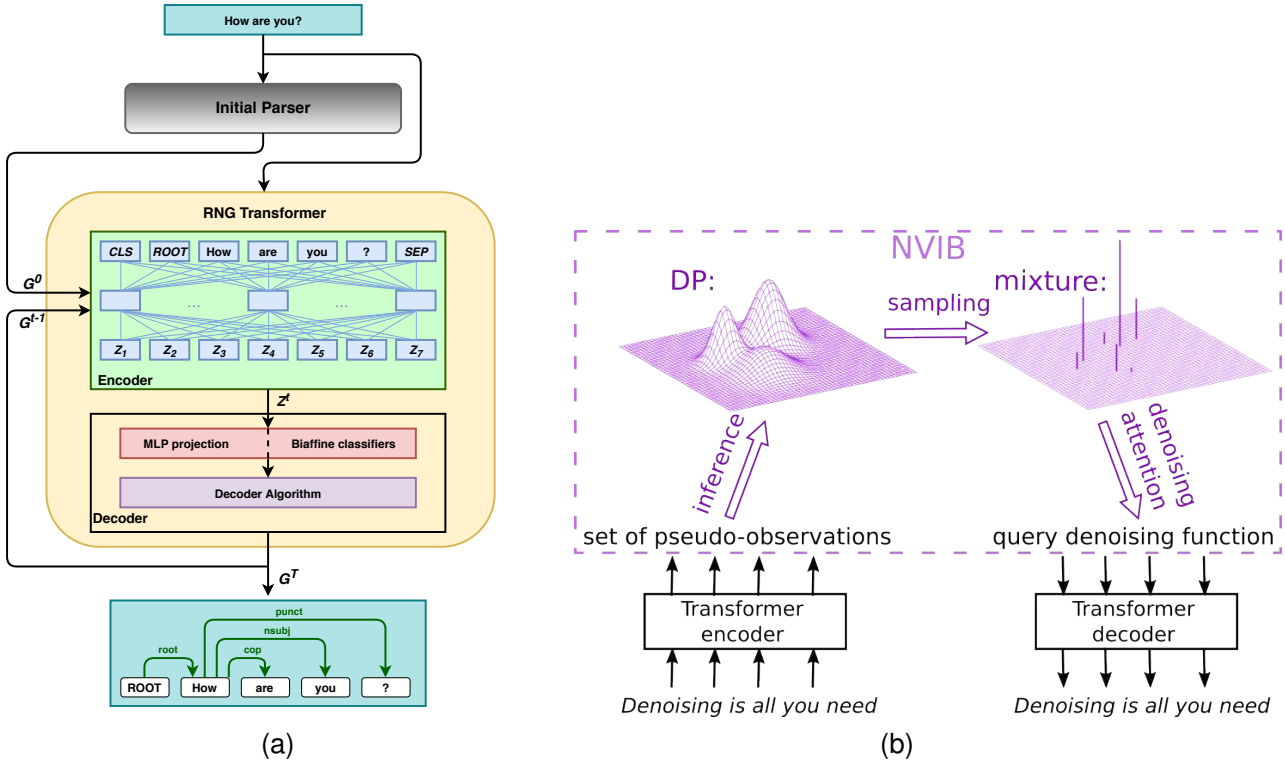


Figure 10: (a) Recursive Non-autoregressive Graph2Graph Transformer for syntactic parsing. (b) Nonparametric Variational Auto-Encoder for inducing varying-sized text embeddings.

our work on inducing what vectors should be included in the latent representation. Our work on using Slot-Attention for this problem is under review (see 2021 arxiv paper), and our work on Bayesian nonparametrics for this problem was released as an arxiv paper and is also under review. This work on variational Bayesian nonparametrics for learning attention-based representations (Figure 10(b)) provides both fundamental insights into the nature of Transformers' latent representations (as nonparametric mixture distributions), and a practical Variational Information Bottleneck (NVIB) regulariser for Transformers.

Key publications

[1] Lesly Miculicich and James Henderson. "Graph Refinement for Coreference Resolution". In *Findings of Association for Computational Linguistics: ACL 2022*, 2022.

[2] Florian Mai and James Henderson. "Bag-of-Vectors Autoencoders for Unsupervised Conditional Text Generation". In *Proc. 2nd Conference of the Asia-Pacific Chapter of the ACL*, 2022.

[3] Andreas Marfurt, Ashley Thornton, David Sylvan, Lonneke van der Plas and James Henderson. "A Corpus and Evaluation for Predicting Semi-Structured Human Annotations". In *Workshop on Generation, Evaluation and Metrics (GEM)*, 2022.

[4] Andreas Marfurt and James Henderson. "Unsupervised Token-level Hallucination Detection from Summary Generation By-products". In *Workshop on Generation, Evaluation and Metrics* , 2022.

[5] Rabeeh Karimi Mahabadi, Luke Zettlemoyer, James Henderson, Saeidi Marzieh, Lambert Mathias, Veselin Stoyanov and Majid Yazdani. "PERFECT: Prompt-free and Efficient Few-shot Learning with Language Models". In *Proc. 60th Annual Meeting of the ACL*, 2022.

[6] Florian Mai, Arnaud Pannatier, Fabio Fehr, Haolin Chen, François Marelli, Francois Fleuret and James Henderson. "HyperMixer: An MLP-based Green AI Alternative to Transformers". In *arxiv*, 2022.

[7] James Henderson and Fabio Fehr. "A Variational AutoEncoder for Transformers with Nonparametric Variational Information Bottleneck". In *arxiv*, 2022.

3.8 Computational Bioimaging

Head: Prof. Michael Liebling (MS, EPFL, 2000; PhD, EPFL 2004; postdoc, Caltech, 2004–2007; Assistant Prof (2007–2013), Associate Prof (2013–2017), Associate Adjunct Prof (2017–) UC Santa Barbara (UCSB))

Group overview: Research in the Computational Bioimaging Group focuses on developing image acquisition, reconstruction and analysis algorithms to study live biological systems. Practical tools aim at (i) extending the physical limits of imaging hardware via techniques including super-resolution and deconvolution methods, and (ii) quantitative analysis of complex biological systems: motion-based image analysis, cell tracking, microscopic fluid flow estimation, and integration of multi-modality images.

In 2022, the Computational Bioimaging Group was composed of the head of the group, one PhD students, a junior developer and an intern.

Key scientific outputs: Recent milestones include temporal super-resolution methods for sensitive fluorescence and thermal cameras and approaches for quantitative heart development imaging in animal models.

Additional information and a list of projects are available at www.idiap.ch/cbi.

Post-acquisition Frame Rate Improvement for Multi-channel Fluorescence Cardiac Microscopy

Optical fluorescence microscopy is frequently used to image the development of organs within animal models. The developing heart is a particularly challenging organ to observe, as the fast contractions and small size require particularly high frame rates. As the fluorescence signal is weak, it requires sensitive cameras, whose often limited frame rate is insufficient to capture the fastest motions of the heart. Furthermore the problem is even more challenging when observing multiple fluorophores. To address these challenges, we have proposed a method to virtually increase the frame rate of image series of the beating heart, which can be implemented with low frame rate cameras by switching illumination according to predefined patterns and collect images over multiple heartbeats and multiple wavelength channels. For implementation, we use a light-sheet microscope with a custom-built illumination controller that is synchronized with the camera. We propose an illumination protocol that alternates between low-intensity bright-field signals and high-intensity laser signals. We combine the acquired images in a post-processing step that sorts the sequences using the bright-field signal as a common reference. We carried out simulation experiments on a multi-channel synthetic phantom of a beating heart. Our experiments show that the frame rate can be virtually increased with a performance that depends on the number of reference images and the noise level at which they can be acquired. Our proposed solution both solves the problem of insufficient frame rate by virtually increasing it, as well as the temporal registration of uncorrelated multi-channel image sequences. The observed performance increase is sufficient to observe motions at speeds typically encountered in cardiac imaging. We foresee that this method can be implemented in a wide variety of existing multi-channel imaging problems of repeating phenomena observed under controlled lighting.

Local estimation of parametric point spread functions in thermal images via convolutional neural networks

Thermal image formation can be modeled as the convolution of an ideal image with a point spread function (PSF) that characterizes the optical degradations. Although simple space-invariant models are sufficient to model diffraction-limited optical systems, they cannot capture local variations that arise because of nonuniform blur. Such degradations are common when the depth of field is limited or when the scene involves motion. Although space-variant deconvolution methods exist, they often require knowledge of the local PSF. We adapt a local PSF estimation method (based on a learning approach and initially designed for visible light microscopy) to thermal images. The architecture of our model uses a ResNet-34 convolutional

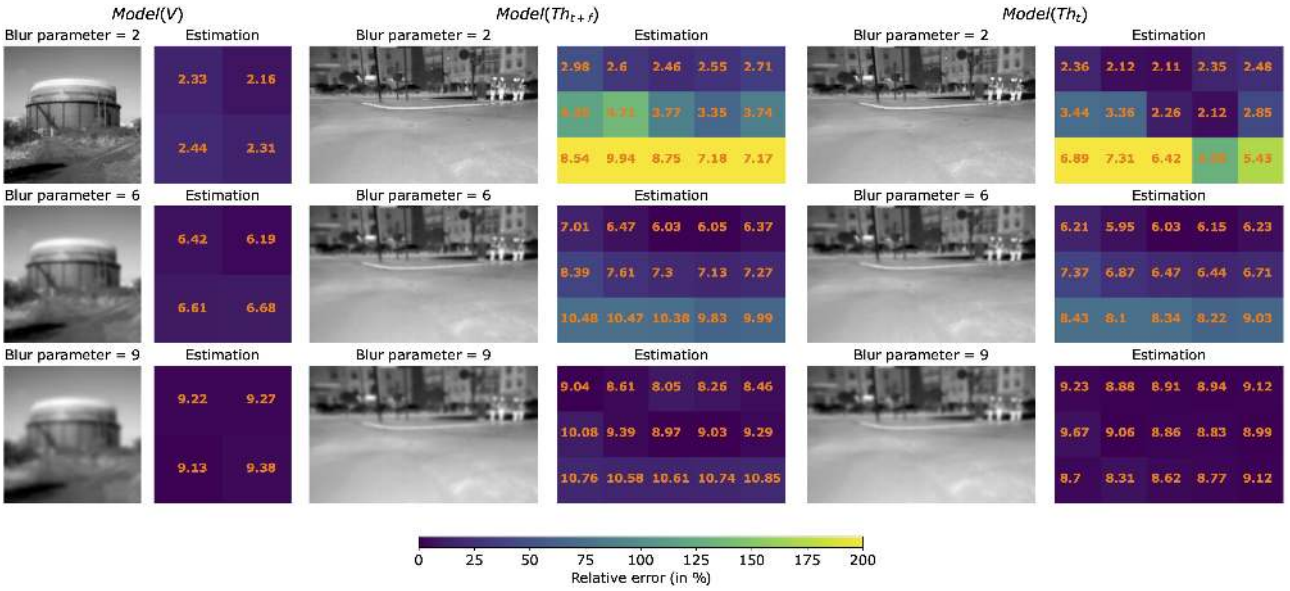


Figure 11: Comparison of the local estimation of the PSF parameter, in areas of size 128 by 128 pixels, The test set is of the same modality as the training set, from left to right: visible images, thermal images containing both textured and flat areas, thermal images containing only textured areas). The example images were blurred with a Gaussian of full width half maximum equal to 2, 6 or 9 pixels. To the right of each blurred image, the map of estimated parameters is shown, where the numbers correspond to the estimated parameter in each region and the background colour represents relative error in that area. (Figure adapted from Piras et al. Proc. SPIE 2022)

neural network (CNN) that we trained on a large thermal image data set (CVC-14) that we split in training, tuning, and evaluation subsets. We annotated the sets by synthetically blurring sharp patches in the images with PSFs whose parameters covered a range of values, thereby producing pairs of sharp and blurred images, which could be used for supervised training and ground truth evaluation. We observe that our method is efficient at recovering PSFs when their width is larger than the size of a pixel. The estimation accuracy depends on the careful selection of training images that contain a wide range of spatial frequencies. In conclusion, while local PSF parameter estimation via a trained CNN can be efficient and versatile, it requires selecting a large and varied training data set. Local deconvolution methods for thermal images could benefit from our proposed PSF estimation method.

Key publications

[1] Y. Liu, J. Dong, T. Pham, F. Marelli, and M. Unser, “Mechanical artifacts in optical projection tomography: classification and automatic calibration,” Opt. Continuum 1, 2577-2589 (2022)

[2] F. Piras, E. De Moura Presa, P. Wellig and M. Liebling, “Local estimation of parametric point spread functions in thermal images via convolutional neural networks,” in: Proc. SPIE 12270, Target and Background Signatures VIII, 1227009, 2022

3.9 Biometrics Security and Privacy

Head: Prof. Sébastien Marcel (PhD, University of Rennes, France, 2000; Visiting Professor, University of Cagliari, 2010; Lecturer, EPFL since 2013; Lecturer, UNIL since 2018; Lecturer, Master AI since 2019, IEEE Senior Member)

Group overview: Biometrics refers to the automatic recognition of individuals based on their physiological and/or behavioural characteristics. The Biometrics Security and Privacy (BSP) group at Idiap focuses on four main areas of research:

- **Biometric recognition.** We investigate the development of accurate and fair (unbiased) recognition algorithms, notably for face, voice, and vein biometric modalities.
- **Presentation attack detection (PAD).** We look for new and better ways of detecting presentation attacks (direct attacks) on face, voice, and vein biometric recognition systems.
- **Morphing attack detection and deepfake detection.** We investigate the threat of morphing attacks (indirect attacks) and more generally *deepfakes* to develop effective morphing and deepfake detection techniques.
- **Biometric template protection.** We research effective methods of preserving both the security of biometric recognition systems and the privacy of their users by protecting the biometric models ("templates") that are employed by the system for recognition purposes.

The BSP group prioritises reproducibility in research. This is important for ensuring that our work can be both verified and built upon by the wider research community. To enable reproducibility, we mainly make use of our Python-based signal-processing and machine-learning toolbox, Bob (http://www.idiap.ch/software/bob/), which we make freely available for academic purposes. The group also develops and maintains the BEAT platform (https://www.beat-eu.org/platform/), an *MLaaS* platform compliant with Swiss and European data-security norms.

The group participates in several large-scale biometrics projects at Swiss (SNSF, Hasler), European (H2020) or world-wide levels (eg. IARPA/DARPA) but also conducts projects directly with companies (IDEMIA, SICPA, Huawei, Google, . . .).

The BSP group provides also expertise to the Swiss Center for Biometrics Research and Testing, which, among other things, carries out tests and evaluations on commercial products related to biometrics.

The BSP group in 2022 was composed of 1 head of group, 7 research associates, 2 postdocs, 4 PhD students, 2 interns.

Key scientific outputs: The BSP group has been pioneering the work on mobile biometrics (face and speaker recognition) and on PAD in face and speaker recognition by sharing the first open databases, organising the first International competitions and producing the first reproducible research studies in the domain. Regarding face PAD, the group confirmed that the current trend using discriminant classifiers is prone to over-fitting hence resulting in a lack of generalisation on unseen presentation attacks. These results question the efficiency and practicality of the existing PAD systems, as well as, call for creation of databases with larger variety of realistic presentation attacks. The BSP group also investigated approaches for heterogeneous face recognition, vein recognition and more recently deepfakes to assess its threat to face recognition. More recently, the BSP group proposed a new scheme to evaluate the fairness of AI-driven face recognition systems.

Additional information and a list of projects are available from www.idiap.ch/biometric.

Deepfakes: a New Threat to Face Recognition?

It is becoming increasingly easy to automatically replace a face of one person in a video with the face of another person by using a pre-trained generative adversarial network (GAN). Recent public scandals call for automated ways to detect these deepfake videos. To help developing such methods, we produced and analysed deepfake videos. We showed that the state of the art face recognition systems based on Neural Networks are vulnerable to deepfake videos, with about 90% false acceptance rates , which means methods for detecting deepfake videos are necessary. By considering several baseline approaches, we found that audio-visual approach based on lip-sync inconsistency detection was not able to distinguish deepfake videos. The best performing method, which is based on visual quality metrics and is often used in presentation attack detection domain, resulted in 8.97% equal error rate on high quality deepfakes. Our experiments demonstrate that GAN-generated deepfake videos are challenging for both face recognition systems and existing detection methods, and the further development of face swapping technology will make it even more so.

Figure 12: Illustration of deepfakes

Presentation attack detection

One important aspect of biometric systems is their reliability not only when assaulted by impostors, but also under different types of attacks. One possible security treat is presentation attacks (aka spoofing attacks): an action of outwitting a biometric sensor by presenting a counterfeit biometric evidence of a valid user. It is a direct attack to the sensory input of the biometric system and the attacker does not need previous knowledge about the recognition algorithm. Most of the biometric modalities are not resistant to presentation attacks: a biometric system is usually designed to only recognise identities without concern whether the sample comes from a live person or not. Despite the existence of very sophisticated biometric systems nowadays, the task of implementing presentation attack detection (PAD aka anti-spoofing) schemes for them has attracted much less attention.

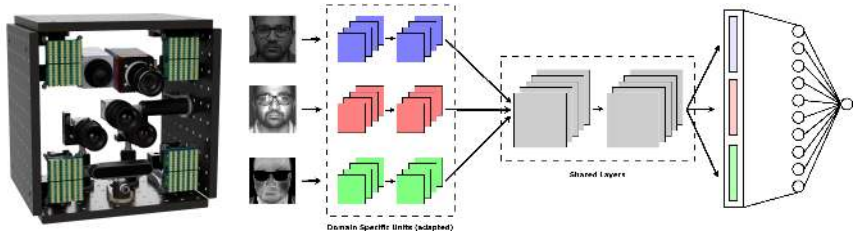


Figure 13: Illustration of multi-spectral face Presentation Attack Detection (PAD) device (left) and architecture (right).

We developed a multi-spectral face device (Figure 13) to capture synchronised, high speed and high resolution image sequences under different image domains (VIS, NIR, SWIR, Thermal and 3D). The main

hypothesis is that bona fide samples are easier to discriminate from presentation attacks with the appropriate combination of image domains. We proposed a novel deep convolutional neural network architecture (Figure 13) to learn multi-spectral complementary information.

Heterogeneous face recognition

The task of Heterogeneous Face Recognition (HFR, Figure 14) consists in to match face images that were sensed in different modalities, such as sketches to photographs, thermal images to photographs or near infrared to photographs. We demonstrated that high level features of deep convolutional neural networks trained on visual spectra images are domain independent and can be used to encode faces sensed in different image domains. We also proposed a surprisingly simple, yet, very effective method for matching face images across different sensing modalities. The core idea of this approach is to add a novel neural network block called Prepended Domain Transformer (PDT) in front of a pre-trained face recognition (FR) model to address the domain gap. Retraining this new block with few paired samples in a contrastive learning setup was enough to achieve state-of-the-art performance in many HFR benchmarks.

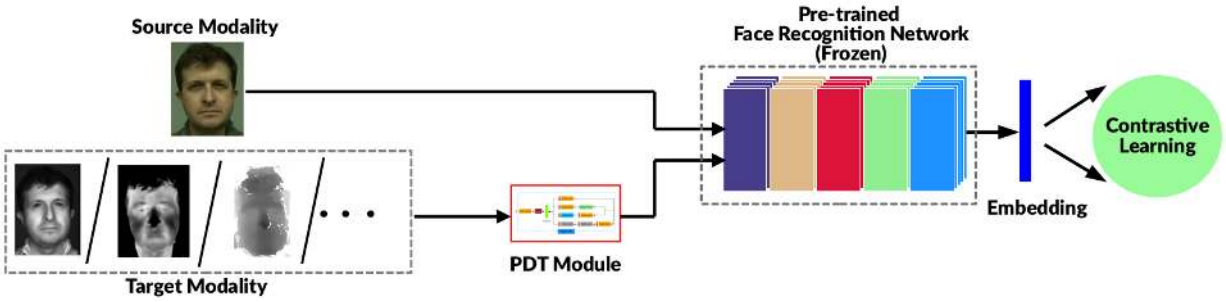


Figure 14: Target domain images are transformed using the proposed Prepended Domain Transformer (PDT) block to a given frozen FR model.

Towards privacy-enhanced face recognition: inversion of deep face templates

State-of-the-art face recognition (FR) systems generally use deep convolutional neural networks (CNNs) to extract deep features, called embeddings, from face images. The face embeddings are stored in the system's database and are used for recognition of the enrolled system users. Hence, these features convey important information about the user's identity, and therefore any attack using the face embeddings jeopardizes the user's security and privacy. We proposed a CNN-based structure to reconstruct face images from face embeddings (Figure 15) and trained with a multi-term loss function.

Swiss Center for Biometrics Research and Testing

In 2014, the Idiap Research Institute launched the "Swiss Centre for Biometrics Research and Testing" (www.biometrics-center.ch), a competence centre within the Institute following recent successes in coordinating International research projects in Biometrics (MOBIO, TABULA RASA and BEAT). The aim of this centre is to serve as a legacy for these projects and to push for industry-driven research and testing in biometrics.

The centre attracted the attention of large companies (license, research and testing agreements) and led to many new projects (DARPA, IARPA, InnoSuisse). In 2020, the centre has developed over three directions:

- Maintaining and evolving the BEAT platform: The platform is now used in multiple research projects such as ALLIES, LEARN-REAL and the major H2020 AI4EU project. In parallel we are working towards improving the platform with the aim to create the Idiap AI platform.

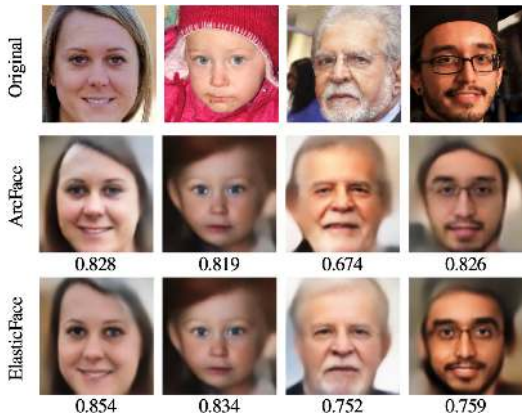


Figure 15: Sample face images (first row) and their corresponding reconstructed face images from ArcFace (second row) and ElasticFace (third row) embeddings. The values indicate cosine similarity between the original and reconstructed image embeddings. The decision thresholds corresponding to FMR $= 10^{-3}$ are 0.37 and 0.41 for ArcFace and ElasticFace, respectively.

- Evaluation and Testing: we engaged with the FIDO Alliance (https://fidoalliance.org), focused on providing open and free authentication standards to help reduce the world's reliance on passwords, and became an accredited FIDO lab able to perform certification of biometrics products. Since 2020 we are also an Android accredited lab and we conducted several evaluations of biometric products.
- Joining the CITeR cooperative research center: we joined the US CITeR (Center for Identification Technology Research) National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) as the first non-US academic site with two affiliates (IDEMIA and SCIPA) to fund our research activities.

Key publications

[1] A. George, A. Mohammadi and S. Marcel, "Prepended Domain Transformer: Heterogeneous Face Recognition without Bells and Whistles", *IEEE Transactions on Information Forensics and Security (TIFS)*, 2022

[2] K. Kotwal, S. Bhattacharjee, P. Abbet, Z. Mostaani, H. Wei, X. Wenkang, Z. Yaxi and S. Marcel, "Domain-Specific Adaptation of CNN for Detecting Face Presentation Attacks in NIR" *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 2022.

[3] V. Krivokuca and S.Marcel, "Biometric Template Protection for Neural-Network-based Face Recognition Systems: A Survey of Methods and Evaluation Techniques", *IEEE Transactions on Information Forensics and Security (TIFS)*, 2022.

[4] V. Krivokuca and S.Marcel, "Towards Protecting Face Embeddings in Mobile Face Verification Scenarios", *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 2022.

[5] H. Otroshi, V. Krivokuca and S. Marcel, "Face Reconstruction from Deep Facial Embeddings using a Convolutional Neural Network". *IEEE International Conference on Image Processing (ICIP)*, 2022

[6] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, "Are GAN-based Morphs Threatening Face Recognition?", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2022.

[7] A. George and S. Marcel, "Cross Modal Focal Loss for RGBD Face Anti-Spoofing", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

3.10 Biosignal Processing

Head: Dr. André Anjos (BSc, MSc & PhD, Federal University of Rio de Janeiro, Brazil, 1999, 2001, 2006; Visiting Professor, State University of Bauru, 2015; Lecturer, EPFL, 2013–; Lecturer, Master of AI, 2019–)

Group overview: The Biosignal Processing group sits in the intersection between AI and Healthcare. It studies the use of Machine Learning for computer-aided diagnosis, prognosis, decision support, and risk analysis from a variety of bio-signals, including, but not limited to, imaging, clinical, demographical, and temporal data. The group has a special focus in reproducibility and the responsible use of AI, with an interest on fairness, explainability, and safety. It leverages on Idiap's expertise on human subject handling, data acquisition, open science and data processing.

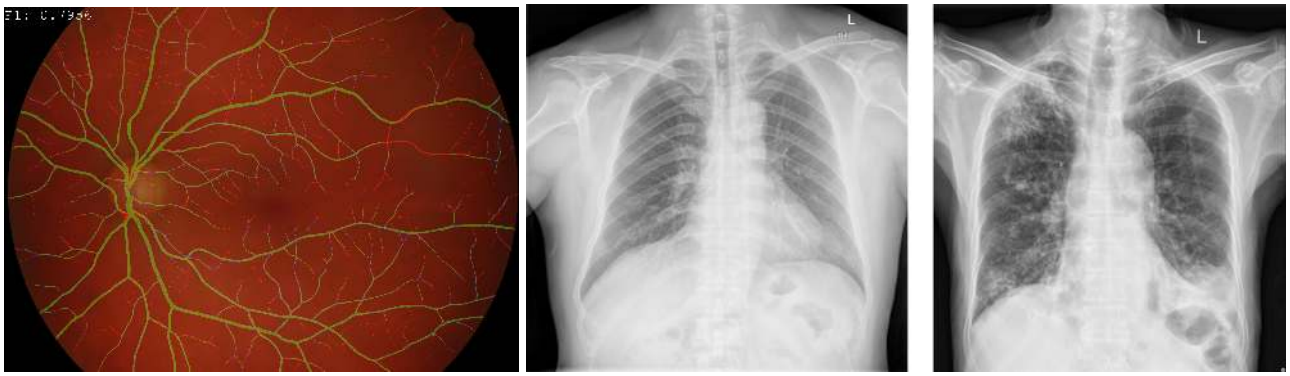
In 2022, the Biosignal Processing Group was composed of the head of the group, a Postdoc researcher, a Ph.D. student, 3 master students, and 2 research interns. The group has active projects with hospital centres in Switzerland, France, and Brazil.

Key scientific outputs: The group currently develops 4 thematic areas of research at the intersection between Artificial Intelligence and Medicine: computer-aided diagnosis (CAD) and decision support from imaging data (e.g. ophthalmic, radiography, histopathology), computer-aided risk prediction, the analysis of time sequences (e.g. vital signs or electroencephalo/cardiography), and, finally, reproducibility in AI research. We are particularly interested in the responsible use (explainability, fairness) of AI models we build, as well as challenges associated to data scarcity in the medical domain. The milestones for 2022 include continued work in above cited areas, and the development of new projects on computed tomography (CT) for the detection of rare diseases, further projects in Ophalmology, reproducibility, and, finally, continued participation in EPFL courses and the Idiap's AI master program. Further funding was requested to the Swiss National Science Foundation to expand the team.

Additional information and a list of projects are available from www.idiap.ch/bio-sig.

Semantic Segmentation for Medical Imaging
Since the introduction of U-Nets in 2015, the field of medical image segmentation has seen renewed interest bringing in a variety of fully convolutional (deep) neural network (FCN) architectures for binary and multi-class segmentation problems promising very attractive results, with applications in computed tomography, retinography, and histopathology to cite a few. Despite the incredible progress, the lack of annotated images (due to cost), and rigor in the comparison of trained models has led the community to believe larger and more dense networks provide better results. We addressed these gaps in different ways. The first was to conduct and publish rigorous (open source, reproducible) benchmarks with popular retinography datasets and state-of-the-art FCN models in which we throw a new light over some published figures, opening space for new developments. In addition, we proposed a simple extension, of a compact U-Net architecture, dubbed W-Net, by concatenating two U-Nets together, which reaches outstanding performance on several popular datasets, still using orders of magnitude less learnable weights than any previously published approach.

Computer-aided Diagnosis for Tuberculosis
Tuberculosis (TB) is one of the leading causes of death from a single infectious agent. In many high-burden regions around the world, which often lack specialized healthcare professionals, Chest X-Ray (CXR) exams continue to be of vital importance in the diagnosis and follow-up of the various presentations of the disease. In this context, we investigate the benefits of automatic Pulmonary Tuberculosis (PTB) detection methods based on radiological signs found on CXR. Contrary to direct scoring from images, implemented in most related work, indirect detection offers natural interpretability of automated reasoning. We identify generalization difficulties for direct detection models trained exclusively on the modest amount of publicly available CXR images from PTB patients. We subsequently show that a model, pre-trained on tens of thousands of CXR images using automatically annotated radiological signs, offers a more adequate base for development. By relaying radiological signs through a simple linear classifier, one is able to obtain



(a) Vessel segmentation for Retinography – predicted vessel maps vs. ground truths for a DL model evaluated on the high-resolution HRF test-set. True positives, false positives and false negatives are displayed in green, blue and red respectively.

(b) Radiological signs on healthy (left) and PTB-affected lungs (right).

state-of-the-art results on all three publicly available datasets. We further discuss limitations imposed by the limited number of PTB-specific radiological signs available on public datasets, and evaluate possible performance gains that could be obtained if more were available. This work is fully reproducible.

Reproducible Research

Since the last decade, we have been actively looking at the reproducibility of published work and how to lower the entrance barrier of publication readers. We argue it is insufficient, in most cases, to only publish software leading to results if original data remains inaccessible. Reproducibility should imply in the following characteristics: repeatability, share-ability, extensibility and stability, which is not guaranteed by most published material to date. We propose a software suite called Bob[32] that possesses such characteristics, demonstrating its flexibility to various tasks. From another perspective, there are legitimate cases in which raw data leading to research conclusions cannot be published. Furthermore, in a growing number of use-cases, the availability of both software does not translate to an accessible reproducibility scenario. To bridge this gap, we built an open platform for research[33] in computational sciences related to pattern recognition and machine learning, to help on the development, reproducibility and certification of results obtained in the field.

Key publications

[1] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, and I. Ben Ayed. “State-of-the-art retinal vessel segmentation with minimalistic models”. Nature Scientific Reports, 12(1):6174, April 2022. Number: 1 Publisher: Nature Publishing Group. DOI: 10.1038/s41598-022-09675-y.

[2] G. Raposo, A. Trajman, and A. Anjos. “Pulmonary tuberculosis screening from radiological signs on chest x-ray images using deep models”. In Union World Conference on Lung Health. The Union, November 2022.

[3] M. A. Renzo, N. Fernandez, A. A. Baceti, N. N. Moura Junior, and A. Anjos. “Development of a lung segmentation algorithm for analog imaged chest x-ray: preliminary results”, XV Brazilian Congress on Computational Intelligence. 2021.

[4] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel, “Continuously Reproducing Toolchains in Pattern Recognition and Machine Learning Experiments”, at Thirty-fourth International Conference on Machine Learning, 2017.

[32] https://www.idiap.ch/software/bob
[33] https://www.idiap.ch/software/beat

3.11 Energy Informatics

Head: Dr. Jérôme Kämpf (BSc, University of Kent at Canterbury; MSc, University of Lausanne, 2001 and 2003; PhD, Ecole Polytechnique Fédérale de Lausanne, 2009)

Group overview: The Energy Informatics concepts are to exploit state-of-the-art Information and Communication Technologies to tackle global warming and climate change challenges. Our aim is to increase the integration of renewable and distributed energy sources by making energy systems smarter, and to increase energy efficiency beyond what improvements at component level can achieve. In that vein, the Energy Informatics Group at Idiap researches into ways of simulating energy transition pathways with intelligent control and adjustment mechanisms of evolving shelters with retrofitting and use, renewable energy production and energy storage in a changing climate. Two key application areas (see §2.2) are more deeply studied: the building automation with its adaptation to human behavior and the energy management with its indirect modeling of human activity to anticipate energy needs.

In 2022, the Energy Informatics Group was composed of 1 head of group, 2 postdocs, 3 exchange PhD students, 1 exchange MSc student and 1 intern.

Key scientific outputs: In building automation, surrogate models were developed to evaluate in a blink two comfort-related KPIs: perceived glare and work plane illuminance. The corresponding machine learning models control in quasi real-time blinds and electric lighting in two offices of our Idiap building. Through a longitudinal study with users we demonstrated savings of around 50% in electricity for lighting with dimmable LED without degrading the comfort of users.
In energy management, a Python-based District Heating Network simulation solver named PyDHN was developed and wrapped in a Graphical User Interface named Eguzki 2.0 for its testing by practitioners. Comparisons with monitoring data taken on a partner networks and other similar software showed a good agreement. To ensure replicability, the software and its validation procedure will be published open-source.

Additional information and a list of projects are available from http://www.idiap.ch/energy.

Building automation

Building occupants tend to negatively perceive building automation as it may alter their comfort for the sake of energy savings. The main challenge identified is therefore to control the building infrastructure in order to minimise the energy intake while maintaining a sufficient comfort for the occupants. Noteworthy, human comfort has multiple facets and depends on each individual's history and preferences. As such, the use of smart sensors is a unique opportunity to learn and adapt the automation to the users.

Through the project LUCIDELES, we study the trade-off between energy and visual comfort in buildings. Machine learning based surrogate models were developed to simulate the daylight penetration through Venetian blinds and its effect on work-plane illuminance and glare perceived by the occupants. These predictive models were then used to control optimally the blinds' state to avoid glare while maximising daylight in buildings, thus minimising the need for electric lighting. Two offices labelled 103 and 107.1 were equipped with the control system. In March and April 2022, we ran a longitudinal study with regular users of both rooms, for which entry, daily and exit questionnaires were considered. The results demonstrated 49% and 67% of electricity consumption reduction with respectively 300 lx and 150 lx on the workplane conditions. The control strategy allowed for a reduction of energy use, while creating an interior environment that does not represent discomfort for the occupants. The two rooms continuously run the system and serve as demonstrator for our visitors.

Through the project LUCIDELES, Idiap was acting as Expert within the International Energy Agency (IEA) Solar Heating and Cooling (SHC) Task 61 (task61.iea-shc.org) on Integrated Solutions for Daylighting and Electric Lighting: From component to user centered system efficiency.

Energy management

While urban areas cover about 2% of the planet surface, they consume 3/4 of the total planetary resources. According to the Pareto rule, major energy-related efforts should target cities and their inhabitants. The efforts comprises the integration of renewable and distributed energy sources, which in dense urban centers remains a challenge. An efficient operation of existing infrastructures and planning of new ones is needed to mitigate and adapt to climate change.

Through the work of three exchange PhD students, we have studied 1) machine learning pipelines to evaluate the performance of district heating networks based solely on their thermal behaviour, 2) the adequacy between solar photovoltaic production and electricity consumption by clustering buildings in a city center and 3) the concept of zero power building/district using dynamic simulations. In the first study, genetic programming optimised pipelines were found to reproduce the different elements of a district heating network: the needs of substations, the losses in the network and the fire-tube boilers preparing the hot water. In the second study, a GIS-based approach using Shannon's Entropy as a measure of dispersion for the type of use of buildings allowed to find optimal clusters of buildings reducing the grid interactions. In the third study, different simulation time-steps were compared to tend towards the concept of zero power building and district, and assessed using dedicated KPIs accounting for the electricity produced by PV panels and stored in batteries.

Meanwhile in the project Eguzki, we developed in 2022 a physically-based solver for District Heating Networks (DHN) named PyDHN. The tool, written in Python, aims to solve for mass flow, temperature and pressure in the pipes under given boundary conditions at central heating stations and substations. A Graphical User Interface named Eguzki 2.0 was built around PyDHN and shared with the industrial partners to undergo heavy testing. A validation of the software is on-going by an inter-model comparison and verification against real monitoring on a meshed and multi-heating stations case-study.

CREM - Centre de Recherches Energétiques et Municipales

In 2022, we maintained a close scientific collaboration with CREM (www.crem.ch), an actor specialized in the field of energy sustainability in urban areas. CREM has a large network of communes, public utilities and companies making it an essential partner for case-studies and dissemination.

Key publications

[1] C. Basurto, M. Papinutto, M. Colombo, R. Boghetti, K. Reutter, J. Nembrini, and J. H. Kämpf. Integrating Daylight with General and Task Lighting: A Longitudinal in-the-Wild Study in Individual and Open Space Working Areas. Solar Energy Advances 2 (2022): 100027.

[2] M. Papinutto, R. Boghetti, M. Colombo, C. Basurto, K. Reutter, D. Lalanne, J. H. Kämpf, and J. Nembrini. Saving Energy by Maximising Daylight and Minimising the Impact on Occupants: An Automatic Lighting System Approach. Energy and Buildings 268 (August 2022): 112176.

[3] G. Cillari,F. Fantozzi, A. Franco, and J. Kaempf. Towards Energy Hubs: An Innovative Geographic Information System Based Approach for Cluster Definition. In ICREC 2022 Conference Proceedings, edited by Energy Reports. Elsevier, 2022.

[4] M. Tognoli, 2022. Physical/data-driven dynamic modelling of fire-tube boilers and demand prediction aiming at adaptive optimization of the supply set-point condition (PhD). Politecnico di Milano, Milano.

[5] A. Demir Dilsiz, K. Ng, J. Kämpf, Z. Nagy, 2022. Ranking parameters in urban energy models for various building forms and climates using sensitivity analysis. Building Simulation. DOI: 10.1007/s12273-022-0961-5

3.12 Computation, Cognition, and Language

Head: Prof. Lonneke van der Plas (MPhil Cambidge, UK, 2001; PhD, University of Groningen, The Netherlands, 2008; Junior professor, University of Stuttgart, Germany, since 2012, Associate professor, University of Malta, since 2014)

Group overview: The Computation, Cognition and Language (CCL) group was newly created in 2021. It is composed of the head, Lonneke van der Plas, and one PhD student since September 2022: Molly Petersen. She has already done her candidacy exam on 28.10.2022 and passed at the 1st attempt. We have incorporated a developer from the R&D group for 6 months, who worked on novel concept generation for French. The group will hire a second PhD student (offers about to be sent out), and an intern will start the 1st of April. We also have 2-year funding for a postdoc thanks to Innosuisse funding for which the hoiring process has started. The group aims to build models of language that are inspired by human cognition and linguistic theory thereby working on the frontiers of what current models are capable of.

The group currently focuses on two areas. Firstly, **cross-lingual transfer**: We investigate to what extend language technology tools for one language can be transferred to another language in low-resource scenarios, for example, when transferring a model that automatically determines the sentiment (positive, negative, or neutral) of a sentence in English to a language such as Maltese, for which limited amounts of data are available. Secondly, **modelling creative processes**: We research models that are able to reproduce aspects of human creative processes with the aim of creating technology that supports human creativity.

Key scientific outputs: The group reached several scientific achievements this year. They published three papers. One sheds light on cross-lingual transfer models, more in particular for Maltese, one on the automatic interpretation of compounds based on multimodal models, and one on the evaluation and prediction of semi-structured human annotations, which received the best resources paper award of the GEM workshop. The head-of-group was invited to give eight invited talks in 2022, such as the plenary talk at the SwissMediaForum in Luzern with 300 participants, the talk at the bioTechX conference that had over 2500 visitors, the WBO conference, in Brussels and the AI and Tech Talks at the Digital Law Center of UNIGE. She also became management committee member of two EU COST actions (CA21167 - Universality, diversity and idiosyncrasy in language technology (UniDive), and CA21131 - Enabling multilingual eye-tracking data collection for human and machine language processing research (MultiplEYE)). She became Editor for Phraseology and Multiword Expressions (Language Science Press), and took part in several events such as the International Create Challenge BCVS special edition, which they won, and resulted in a collaboration with the BCVS. .

Pre-training Data Quality and Quantity for a Low-Resource Language

Multilingual language models such as mBERT have seen impressive cross-lingual transfer to a variety of languages, but many low-resource languages remain excluded from these large multilingual language models. In Micallef et al. (2022), we analysed the effect of pre-training with monolingual data for a low-resource language that is not included in mBERT – Maltese – with a range of pre-training set ups. We conducted evaluations with the newly pre-trained models on three morphosyntactic tasks – dependency parsing, part-of-speech tagging, and named-entity recognition – and one semantic classification task – sentiment analysis. We also presented a newly created corpus for Maltese, and determine the effect that the pre-training data size and domain have on the downstream performance. Our results show that using a mixture of pre-training domains is often superior to using Wikipedia text only. We also find that a fraction of this corpus is enough to make significant leaps in performance over Wikipedia-trained models. We pre-train and compare two models on the new corpus: a monolingual BERT model trained from scratch (BERTu), and a further pretrained multilingual BERT (mBERTu). The models achieve state-of-the-art performance on these tasks, despite the new corpus being considerably smaller than typically used corpora for high-resourced languages. On average, BERTu outperforms or performs competitively with mBERTu, and the

largest gains are observed for higher-level tasks.

Visually Grounded Interpretation of Noun-Noun Compounds in English

Noun-noun compounds (NNCs), combinations of nouns that function as a single unit, such as chocolate cake, occur frequently in the English language and are a main vehicle for lexical innovation. New concepts in language are often referred to by means of compounding : time sink, screen time, carbon footprint etc. Accurate NNC interpretation, i.e. determining the implicit relationship between the constituents of a NNC, is crucial for the advancement of many natural language processing tasks, and for understanding creative processes at play in language evolution. Until now, computational NNC interpretation has been limited to approaches involving linguistic representations only. However, much research suggests that grounding linguistic representations in vision or other modalities can increase performance on this and other tasks. In Lang et al. (2022), we present a novel comparison of linguistic and visuo-linguistic representations for the task of NNC interpretation. We frame NNC interpretation as a relation classification task, evaluating on a large, relationally-annotated NNC dataset. We combine distributional word vectors with image vectors to investigate how visual information can help improve NNC interpretation systems. We find that adding visual vectors increases classification performance on our dataset in many cases.

A Corpus and Evaluation for Predicting Semi-Structured Human Annotations

This paper is the result of a cross-disciplinary collaboration between researchers in the social sciences, and the NLU group at Idiap, where the CCL group was involved as well. A wide variety of tasks have been framed as text-to-text tasks to allow processing by sequence-to-sequence models. In Marfurt et al. (2022), we proposed a new task of generating a semi-structured interpretation of a source document. The interpretation is semi-structured in that it contains mandatory and optional fields with free-text information. This structure is surfaced by human annotations, which we standardize and convert to text format. We then propose an evaluation technique that is generally applicable to any such semi-structured annotation, called equivalence classes evaluation. The evaluation technique is efficient and scalable; it creates a large number of evaluation instances from a comparably cheap clustering of the free-text information by domain experts. For our task, we release a dataset about the monetary policy of the Federal Reserve. On this corpus, our evaluation shows larger differences between pretrained models than standard text generation metrics.

Key publications

[1] Kurt Micallef, Albert Gatt, Marc Tanti, Lonneke van der Plas and Claudia Borg. Pre-training Data Quality and Quantity for a Low-Resource Language: New Corpus and BERT Models for Maltese. In Proceedings of the workshop on Deep Learning for Low-Resource NLP. 2022

[2] Inga Lang, Lonneke van der Plas, Malvina Nissim and Albert Gatt. Visually Grounded Interpretation of Noun-Noun Compounds in English. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, Association for Computational Linguistics. 2022

[3] Andreas Marfurt, Ashley Thornton, David Sylvan, Lonneke van der Plas, and James Henderson. A Corpus and Evaluation for Predicting Semi-Structured Human Annotations. In Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM), pages 262–275, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics. 2022

3.13 Reasoning and Explainable AI

Head: Dr. André Freitas (PhD, NUI Galway, Ireland, 2016; Senior Lecturer (Associate Professor), University of Manchester, UK; AI Group Leader, CRUK Manchester Institute, UK)

Group overview:

The Neuro-symbolic AI group aims at developing AI models which are capable of complex, abstract and flexible inference. The group operates at the interface between neural and symbolic AI methods aiming to enable the next generation of explainable, data-efficient and safe AI models. These themes closely dialogue with the need for mitigating some of the risks for the adoption of AI models in real world scenarios. The group is part of Idiap's Cross Research Groups (CRGs) initiative, aiming to promote strategic areas in AI and to facilitate internal integration across research groups, external collaborations and industrial applications.

Key scientific outputs: The group is pioneering the development of neuro-symbolic and natural language inference (NLI) models targeting abstract and scientific domains, including areas such as Cancer Research, Mathematics and Physics. Recent contributions include: the development of novel neuro-symbolic architectures to support better inference control, development of models which merge mechanistic and statistical inference, variational autoencoders and disentanglement for natural language inference, improving the formal understanding of inference and explanatory patterns in scientific discourse, the development of probing, testing and causal frameworks for natural language inference; assessment of the interaction between end-users and AI models, aiming for critical transparency and clinical embedding; development of safe and explainable AI models to support cancer research and industrial applications. Currently, the group focuses on 6 core research streams, which are outlined below.

Building AI models capable of complex abstract inference. *How to develop AI models which can deliver complex, expert-level (e.g. scientific) reasoning and explanations?* In this area, the group focuses on building models capable of encoding complex and abstract inference, with a particular emphasis in the scientific domain. Figure 16 outlines an example of a complex expert-level explanation/argumentation in the context of oncology. Our contributions to this area affirm a balance between the flexibility provided by contemporary deep learning based models and the explicit inference controls delivered by symbolic methods. In linguistic terms, scientific explanations are still poorly defined objects. Our research contributed to define some of the critical components behind scientific inference and explanations, and the recurring linguistic and inference patterns expressed on them. Eliciting these components can inform the construction of models which are capable of abstract inference, better generalisation and which are natively explainable.

Reasoning over natural language & equations. *How to develop AI models which can reason over mathematical text?* A significant part of scientific discourse is expressed as mathematics. The ExplAIn group is currently pioneering new AI methodologies which allows automated inference over mathematical text, where models need to encode both natural language and mathematical expressions. This year, the group established new modeling mechanisms to jointly encode these two symbolic modalities, including modelling the interpretation of variables within mathematical discourse. Our investigation extended to the domain of Physics, where we developed the first corpus for Natural Language Inference over Physics texts.

Safe & controlled inference. *How to develop inference models which balance flexibility, rigour and safety?* We investigated new methods to integrate additional logical constraints into neural-based natural language inference (NLI) models. Our contributions to this area include the analysis and improvement of abstract logical feature modelling within neural NLI models. In order to establish the internal linguistic and inference consistency properties of neural NLI models, we introduced the notion of metamorphic testing in Natural Language Processing (NLP), and systematically applied semantic, inference and causal verification within different domains.

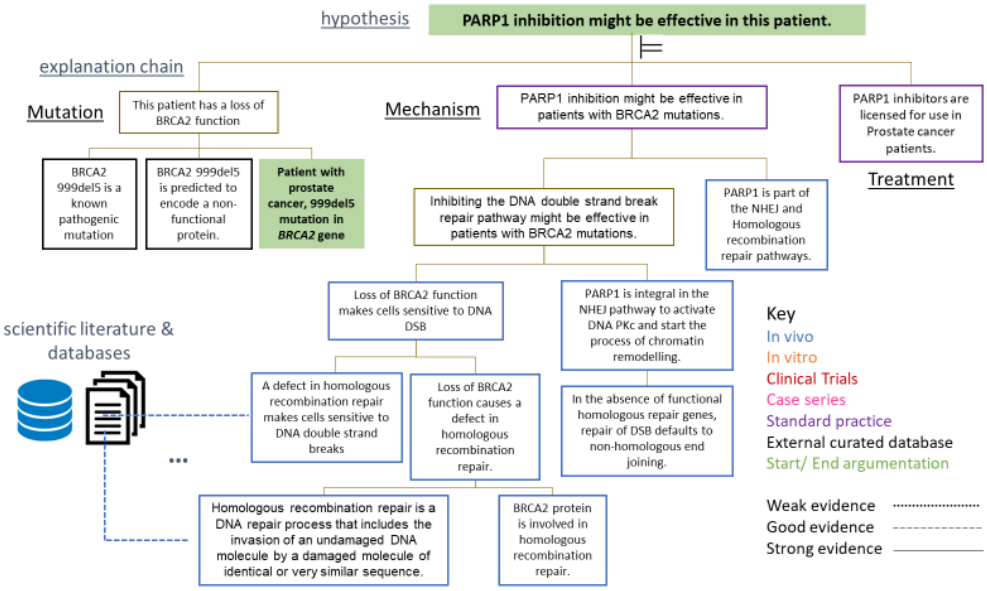


Figure 16: Example of expert-level explanation in the context of oncology.

Transparency & explainability. *How to develop efficient communication mechanisms between end-users and AI models?* The application of machine learning (ML) models in domains such as healthcare requires the introduction of mechanisms which allow end-users to have a critical understanding of the models' underlying inference process. The design of explainabilty mechanisms allows for the communication of these model properties to non-AI experts. During 2021, we pragmatically embedded different standard models of explainability in a clinical decision setting, and found that, despite delivering transparency, these models still lack the usability required by healthcare professionals. The underlying complexity of AI models brings transparency problems even for AI experts. One of the research areas investigated by the group is on how to improve the efficiency of the communication between experts using optimised diagrammatic representations of neural network based models.

Clinical applications & Industrial Collaborations. The group collaborates closely with a major cancer research centre in the UK (CRUK Manchester Institute), on the development of AI models in the oncology domain. Areas of collaboration involve the development of AI models to support treatment recommendation, toxicity prediction and the development of novel biomarkers, with an emphasis on safety, explainability and ethical awareness (SEE). Current projects include the use of natural language inference to support cancer clinical trials (CCTs) and the development of biologically-informed deep learning models for optimising CCTs. Current industrial collaborations involve the use of natural language inference to support the discovery of new antibiotic substances (in partnership with Inflamalps), understanding social trends (in partnership with Bloom) and a cross research group collaboration on Safety and Explainable AI for decision support in orthopaedics (in collaboration with the Biosignal Processing Group and Med4CAST).

Key publications

[1] Julia Rozanova, Marco Valentino, Lucas Cordeiro, André Freitas. "On Interventional Probing in High Dimensions: An NLI Case Study". In *In Findings of the EACL*, 2023.

[2] Danilo S. Carvalho, Giangiacomo Mercatali, Yingji Zhang, André Freitas. "Learning Disentangled Representations for Natural Language Definitions" In *Findings of the EACL*, 2023.

[3] Mokanarangan Thayaparan, Marco Valentino, Deborah Ferreira, Julia Rozanova, André Freitas. "Diff-Explainer: Differentiable Convex Optimization for Explainable Multi-hop Inference". In *TACL*, 2022.

[4] Marco Valentino, Mokanarangan Thayaparan, Deborah Ferreira, André Freitas. "Hybrid Autoregressive Solver for Scalable Abductive Natural Language Inference". In *Proc. of AAAI*, 2022.

[5] Oskar Wysocki, Zili Zhou, Paul O'Regan, Deborah Ferreira, Magdalena Wysocka, Donal Landers, André Freitas. "Transformers and the representation of biomedical background knowledge". In *Computational Linguistics*, 2022.

3.14 Signal Processing for Communication

Head: Dr. Ina Kodrasi (MSc, Jacobs University Bremen, Germany, 2010; PhD, Carl von Ossietzky Universität Oldenburg, Germany, 2015)

Group overview:

The Signal Processing for Communication (SPC) group focuses on tackling challenges arising in acoustic communication such as speaking impairments, hearing impairments, and undesired interferences (e.g., additive noise and reverberation). The high-level objectives of the group are to establish novel digital signal processing and pattern recognition approaches for speech, audio, and multi-modal signals to automatically detect speaking and hearing impairments, provide speaking and hearing assistance, and improve the acoustic communication experience in the presence of undesired interferences.

In 2022, the SPC group consisted of two PhD students and one short-term research visitor in the framework of a COST action on active and assisted living.

Key scientific outputs: During 2022, the SPC group focused on developing automatic methods for detecting pathological speech arising due to neurological disorders such as Parkinson's disease or Amyotrophic Lateral Sclerosis. In this context, some of the key scientific contributions were i) incorporating perceptually-motivated phase transformations in automatic pathological speech detection approaches, ii) establishing a novel adversarial-free supervised representation learning framework which can be used beyond automatic pathological speech detection, and iii) establishing clinically-motivated acoustic features for discriminating multiple sub-types of speech pathologies.

Additional information and a list of projects are available at www.idiap.ch/spc.

Phase representations for pathological speech detection. Mainstream deep learning-based pathological speech detection approaches typically rely on processing the magnitude spectrum of the short-time Fourier transform of input signals, while ignoring the phase spectrum. Although considerable insight about the structure of a signal can be obtained from the magnitude spectrum, the phase spectrum also contains inherent structures which are not immediately apparent due to phase discontinuity (see Figure 17). To reveal meaningful phase structures for pathological speech detection, we have proposed to consider alternative representations such as the modified group delay spectrum and the instantaneous frequency spectrum (see Figure 17) [1]. We have shown that pathological cues are present in all considered phase representations. Further, we have shown that using phase representations as complementary features to the magnitude spectrum as in Figure 18 is very beneficial for deep learning-based pathological speech detection, with the combination of the magnitude and instantaneous frequency spectra yielding a high performance. We believe that the results presented in [1] will raise awareness in the research community about the potential of the phase spectrum for pathological speech detection and motivate research into novel architectures that optimally exploit magnitude and phase information.

Adversarial-free supervised representation learning framework. Speech representations which are robust to pathology-unrelated cues such as speaker identity information are advantageous for automatic pathological speech detection [2]. We had previously proposed an approach to learn speaker identity-invariant representations for pathological speech detection based on adversarial training [2]. However, adversarial training can be challenging, unstable, and sensitive to training parameters. To avoid adversarial training, we have recently proposed to learn speaker-identity invariant representations exploiting a feature separation framework relying on mutual information minimization (see Figure 18) [3]. We have shown that the proposed adversarial-free framework successfully learns speaker identity-invariant representations. Further, we have shown that such representations result in a similar pathological speech detection performance as the representations obtained using adversarial training, while the training procedure is more stable and less sensitive to training parameters. Such a training framework can be used in other

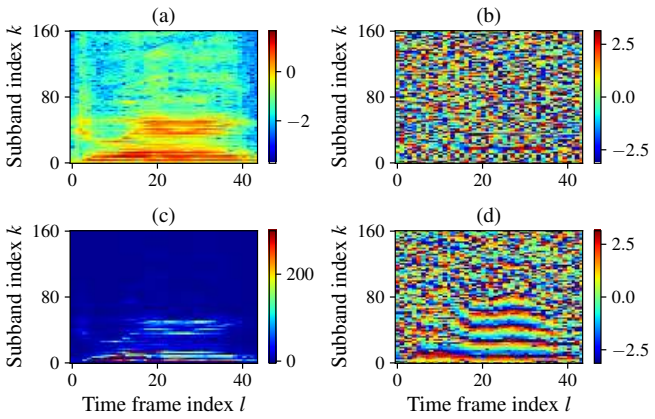


Figure 17: Short-time Fourier transform of an exemplary utterance: (a) magnitude, (b) phase, (c) modified group delay, and (d) instantaneous frequency.

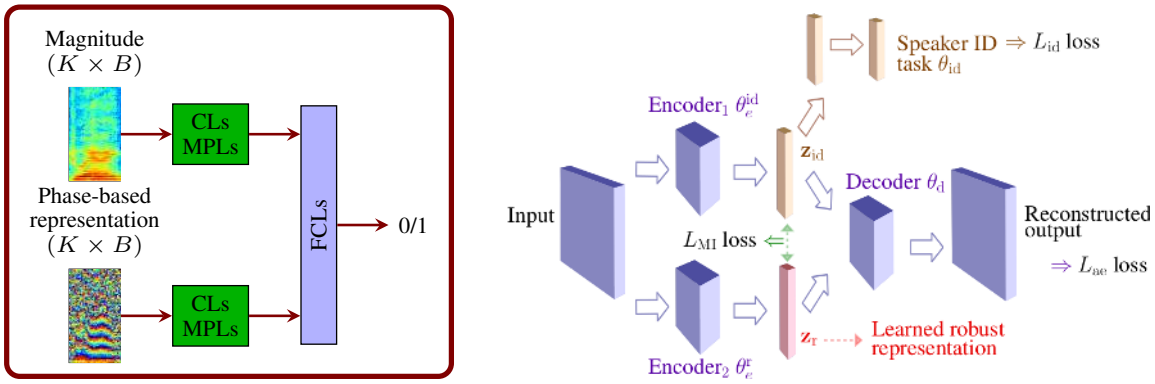


Figure 18: Dual-input convolutional neural network exploiting both magnitude and phase representations (left) and non-adversarial feature separation network (right).

applications aiming to disentangle characteristics of bottleneck representations without using adversarial training.

Clinically-motivated acoustic features for discriminating multiple sub-types of speech pathologies. Using decision trees and the MonPaGe-2.0.s clinical assessment protocol of pathological speech based on a combination of acoustic and perceptive scores, we have shown that acoustically- and perceptually-motivated deviance scores are able to differentiate major sub-types of speech pathologies [4]. Such scores capture deviance on intelligibility, articulation, voice, speech rate, maximum phonation time, prosody, and diadochokinetic rate. In the future, robust methods to automatically extract these scores will be developed.

Key publications

[1] P. Janbakhshi and I. Kodrasi, "Experimental investigation on STFT phase representations for deep learning-based dysarthric speech detection," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Singapore, May 2022, pp. 6477–6481.

[2] P. Janbakhshi and I. Kodrasi, "Supervised speech representation learning for Parkinson's Disease classification," in *Proc. ITG Conference on Speech Communication*, Kiel, Germany, Sept. 2021, pp. 1–5.

[3] P. Janbakhshi and I. Kodrasi, "Adversarial-free speaker identity-invariant representation learning for automatic dysarthric speech classification," in *Proc. Annual Conference of the International Speech Communication Association*, Incheon, Korea, Sept. 2022, pp. 2138–2142.

[4] C, Fougeron, I. Kodrasi, and M. Laganaro, "Differentiation of motor speech disorders through the seven deviance scores from MonPaGe-2.0.s," in Brain Sciences, vol. 12, no. 11, pp. 1471–1487, Oct. 2022.

3.15 Human-centered Robotics and AI

Head: Dr. Emmanuel Senft (MSc, EPFL, Switzerland, 2013; PhD, University of Plymouth, United Kingdom, 2018)

Group overview: The Human-centered Robotics and AI group is one of the new Cross-Research Groups and tackles industrial and societal challenges at the interface of humans and technology. This group's main focus is to develop human-centered robotic systems for end-users to design effective robot behaviors *in situ*; explores how to deploy robots in human environments with high cost of failure, large variability of tasks, and real-world application; and enables robots to be used by anyone regardless of their computer literacy or physical capabilities.

The group was established in September 2022 and Dr. Senft is currently building the team to address the topics outlined above. The first members of the groups will be recruited in 2023 to work on assistive robots for people with neurodegenerative diseases.

Key scientific outputs: Since its inception, the HRAI group has been involved in three publications from previous work. Two were presented at IROS 2022 and one is to be presented at CHI 2023. This last paper presents our initial work on participatory design for robots in assisted living facilities.

The head-of-group has become a program committee member of the IEEE/ACM Human-Robot Interaction conference (HRI), the main venue in the field, and presented the group in a number of invited talks in Kyoto (Japan) and Grenoble (France), and virtually at the Colorado School of Mines.

Additional information and a list of projects are available at www.idiap.ch/hrai.

Assistive Robotics

With the aging of the population and the decrease in the number of caregivers, there is a growing demand for assistant robots to help older adults and people with disabilities or neurodegenerative diseases. These robots can offer valuable assistance by taking care of daily tasks such as fetching items, assisting with mobility, or providing some basic care. In addition to offering valuable assistance to patients, assistive robots can also relieve the pressure on caregivers (formal and informal ones) who may face heavy responsibilities and exhausting working hours. Besides fully autonomous robot, there are also opportunities for teleoperation. For example, a caregiver can use a robot to communicate with a patient, check on their health, and do simple tasks by controlling the robot remotely. By using robot assistants, caregivers and caregivers can devote more time and energy to providing quality care and tending to their own needs. Similarly, robots can also be used directly by patients to regain some autonomy.

Nevertheless the actual role of the robot, the tasks it should do, as well as the interfaces used by the different actors (patients, families, care staff) are not fixed and can vary from one situation to the other.

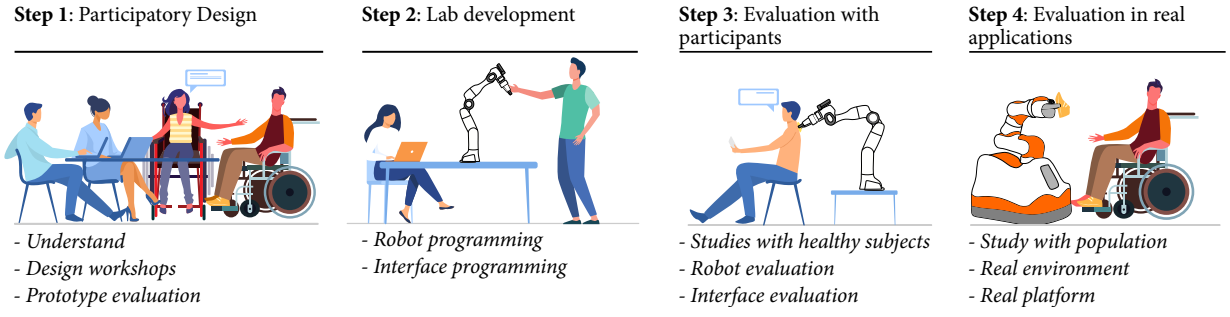


Figure 19: Research method to design assistive robotic systems. [This image has been designed using assets from Freepik.com]

Consequently, it is key to involve the users of such technologies early in the development process. We use multiple Participatory Design methods to co-design these aspects of the interaction with the different actors of the interaction. Figure 19 shows our approach when designing technologies: first we work with the users to define the interaction and interfaces, then we implement these systems, evaluate them in lab studies, and finally deploy the systems in real environments with the real population.

Developing such systems requires solving a large number of challenges. In addition to classic robot technologies (e.g., navigation, object detection, and grasping) human-robot interactions rely on a number of other technology such as effective speech recognition, language processing, and human activity detection. Consequently, this research strongly encourages collaborations with other research groups at Idiap. Additionally, we also explore new interfaces and shared control paradigms to balance robot autonomy and human inputs to achieve useful robot deployments. Finally, as we plan to deploy our robots for extended periods of times, interacting with humans, and receiving inputs from them, we will develop and evaluate new human-in-the-loop learning models.

Key publications

[1] Stegner, L., Senft, E., & Mutlu, B. (2023). "Situated Participatory Design: A Method for In Situ Design of Robotic Interaction with Older Adults." accepted in *Proceedings of the 2023 CHI conference on human factors in computing systems*.

4 Researchers (in addition to permanent staff)

4.1 Research Associates

Nationality	Last name	First name	Supervisor		Start	End (estimated)
CH	BHATTACHARJEE	Sushil	Sébastien	MARCEL	2016-01-01	
DE	BRAUN	Rudolf	Hervé	BOURLARD	2021-03-01	2022-09-30
CH	GEISSBUHLER	David	Sébastien	MARCEL	2018-06-01	
IN	GEORGE	Anjith	Sébastien	MARCEL	2021-01-01	
FR	KOMATY	Alain	Sébastien	MARCEL	2022-06-01	
EE	KORSHUNOV	Pavel	Sébastien	MARCEL	2017-07-01	
IN	KOTWAL	Ketan	Sébastien	MARCEL	2021-08-01	
HR	KRIVOKUĆA HAHN	Vedrana	Sébastien	MARCEL	2021-01-16	
IN	MADIKERI	Srikanth	Petr	MOTLICEK	2018-03-01	
ES	VILLAMIZAR	Michael	Jean-Marc	ODOBEZ	2020-11-01	
MX	VILLATORO TELLO	Esaú	Hervé	BOURLARD	2021-11-01	

4.2 Post-doctoral Scholars

Nationality	Last name	First name	Supervisor		Start	End (estimated)
MX	BASURTO DAVILA	Chantal	Jérôme	KÄMPF	2020-04-01	2022-01-31
AR	BURDISSO	Sergio	Petr	MOTLICEK	2022-02-15	2023-02-15
BR	DE FREITAS PEREIRA	Tiago	Sébastien	MARCEL	2019-03-01	2022-02-28
FR	DELMAS	Maxime	André	FREITAS	2022-10-01	2024-03-31
CH	ECABERT	Christophe	Sébastien	MARCEL	2022-05-01	2023-04-30
MX	JIMÉNEZ DEL TORO	Oscar	André	ANJOS	2022-04-15	2023-04-14
IR	MOHAMMADI	Amir	Sébastien	MARCEL	2020-02-01	2022-01-31
IN	MURALIDHAR	Skanda	Mathew	MAGIMAI DOSS	2022-04-01	2023-03-31
IN	PARIDA	Shantipriya	Petr	MOTLICEK	2019-02-01	2022-01-31
IT	PERONATO	Giuseppe	Jérôme	KÄMPF	2020-09-04	2023-11-30
IT	RACCA	Mattia	Jean-Marc	ODOBEZ	2020-10-01	2022-03-31
IR	RAZEGHI	Behrooz	Sébastien	MARCEL	2022-12-01	2023-10-31
IR	SARFJOO	Saeed	Petr	MOTLICEK	2018-01-01	2022-06-30
CH	SIEGFRIED	Rémy	Jean-Marc	ODOBEZ	2021-10-01	2023-01-15
PT	SILVERIO	João	Sylvain	CALINON	2019-07-01	2022-02-28
IN	SINGH	Muskaan	Petr	MOTLICEK	2022-02-01	2022-10-15
DE	SOUSA EWERTON	Marco	Jean-Marc	ODOBEZ	2019-08-15	2022-04-30
PL	TKACZUK	Jakub	Petr	MOTLICEK	2022-01-15	2023-01-14
CH	TORNAY	Sandrine	Mathew	MAGIMAI DOSS	2021-03-01	2024-08-31
DE	VLASENKO	Bogdan	Mathew	MAGIMAI DOSS	2021-02-15	2023-03-31
CH	ZUFFEREY	Marie	André	FREITAS	2021-09-15	2022-05-31

4.3 PhD Students

Nationality	Last name	First name	Supervisor		Start	End (estimated)
CN	CHEN	Haolin	Philip	GARNER	2021-08-30	2025-08-01
CH	SARKAR	Eklavya	Mathew	MAGIMAI DOSS	2021-03-01	2025-02-28
IR	JOHARI	Seyed M. M.	François	FLEURET	2020-01-01	2023-12-31
CN	HE	Mutian	Philip	GARNER	2022-04-01	2026-03-31
DE	FRITSCH	Julian	Mathew	MAGIMAI DOSS	2018-07-01	2022-06-30
CH	VUILLECARD	Pierre	Jean-Marc	ODOBEZ	2022-06-01	2026-05-31
IR	RAZMJOO FARD	Amirreza	Sylvain	CALINON	2020-12-01	2024-11-30
FR	COURDIER	Evann	François	FLEURET	2019-03-01	2023-02-28
CN	LI	Yiming	Sylvain	CALINON	2022-09-01	2026-08-31
CH	LÖW	Tobias	Sylvain	CALINON	2021-01-04	2024-12-31
CO	ZULUAGA GOMEZ	Juan Pablo	Petr	MOTLICEK	2020-01-01	2023-12-31
BR	GARCIA SCHU PEIXOTO	Guilherme	Ina	KODRASI	2022-01-15	2023-01-31
ID	LEMBONO	Teguh	Sylvain	CALINON	2018-07-23	2022-06-30
IR	SAJADMANESH	Sina	Daniel	GATICA-PEREZ	2019-05-15	2023-05-15
IT	COMAN	Andrei	James	HENDERSON	2020-10-01	2024-09-30
IR	KARIMI MAHABADI	Rabeeh	James	HENDERSON	2018-10-01	2022-09-30
IR	MOSTAANI	Zohreh	Mathew	MAGIMAI DOSS	2020-02-01	2024-01-31
IR	BEHJATI	Melika	James	HENDERSON	2020-09-01	2023-08-31
IN	VYAS	Apoorv	Hervé	BOURLARD	2018-07-15	2022-07-15
TR	GIRGIN	Hakan	Sylvain	CALINON	2018-09-01	2022-08-31
IN	GUPTA	Anshul	Jean-Marc	ODOBEZ	2021-01-01	2024-12-31
CN	XUE	Teng	Sylvain	CALINON	2021-11-01	2025-10-31
DE	MAI	Florian	James	HENDERSON	2018-10-01	2022-09-30
IN	PUROHIT	Tilak	Mathew	MAGIMAI DOSS	2021-06-01	2025-05-31
MA	TAFASCA	Samy	Jean-Marc	ODOBEZ	2021-03-01	2025-02-28
IN	ANNAPUREDDY	Ravinithesh Reddy	Daniel	GATICA-PEREZ	2022-10-01	2026-09-30
US	MATOBA	Kyle	François	FLEURET	2020-08-01	2023-07-31
IN	SHETTY	Suhan	Sylvain	CALINON	2019-06-03	2023-06-03
IR	RAHIMI NOSHANAGH	Parsa	Sébastien	MARCEL	2022-08-01	2026-07-31
US	PETERSEN	Molly	Lonneke	VAN DER PLAS	2022-09-01	2026-08-31
CH	UNNERVIK	Alex	Sébastien	MARCEL	2020-03-01	2024-02-29
BE	MARELLI	François	Michael	LIEBLING	2018-10-01	2022-09-30
IN	SIVAPRASAD	Prabhu	François	FLEURET	2018-11-01	2022-10-31
IR	OTROSHI SHAHREZA	Hatef	Sébastien	MARCEL	2020-03-01	2024-02-29
IN	TARIGOPULA	Neha	Mathew	MAGIMAI DOSS	2021-03-08	2025-02-28
IT	MESSORI	Elisa	Raphaëlle	LUISIER	2022-09-19	2026-09-18
IR	MOHAMMADSHAHI	Alireza	James	HENDERSON	2019-09-01	2022-08-31
LK	MEEGAHAPOLA	lakmal	Daniel	GATICA-PEREZ	2019-06-24	2023-06-23
RU	THORBECKE	Iuliia	Petr	MOTLICEK	2021-01-18	2025-01-18
CH	MARFURT	Andreas	James	HENDERSON	2018-11-01	2022-08-31
CH	PANNATIER	Arnaud	François	FLEURET	2020-03-01	2024-02-29
TR	BILALOGLU	Cem	Sylvain	CALINON	2022-09-01	2026-08-31
SK	FAJČÍK	Martin	Petr	MOTLICEK	2021-11-01	2022-03-14
TR	KABIL	Selen	Hervé	BOURLARD	2018-05-01	2022-04-30
DE	JANKOWSKI	Julius	Sylvain	CALINON	2020-04-01	2024-03-31

DE	HERMANN	Enno	Mathew	MAGIMAI DOSS	2018-07-01	2022-06-30
BE	COPPIETERS DE GIBSON	Louise	Philip	GARNER	2020-10-01	2024-09-30
IT	BOGHETTI	Roberto	Jérôme	KÄMPF	2021-06-01	2022-01-31
IN	PRASAD	Amrutha	Petr	MOTLICEK	2020-08-01	2024-07-31
CH	BITTAR	Alexandre	Philip	GARNER	2020-03-02	2024-03-02
IR	JANBAKHSHI	Parvaneh	Ina	KODRASI	2018-02-01	2022-01-31
CH	FEHR	Fabio	James	HENDERSON	2021-02-01	2025-01-31
CH	COLBOIS	Laurent	Sébastien	MARCEL	2021-01-01	2024-12-31
CN	ZHANG	Yan	Sylvain	CALINON	2022-10-01	2026-09-30

4.4 Interns

Nationality	Last name	First name	Supervisor		Start	End (estimated)
ES	ALONSO DEL BARRIO	David	Daniel	GATICA-PEREZ	2022-01-01	2022-05-31
IT	BILARDO	Matteo	Jérôme	KÄMPF	2022-09-19	2022-12-19
IT	BONTEMPELLI	Andrea	Daniel	GATICA-PEREZ	2022-11-14	2023-01-27
FR	BOUTON-BESSAC	Emma	Daniel	GATICA-PEREZ	2021-09-20	2022-01-31
FR	BRAUD	Lucas	Sébastien	MARCEL	2022-09-20	2023-01-06
FR	BROS	Victor	Daniel	GATICA-PEREZ	2021-02-01	2022-01-31
IT	BRUNINI	Gabriele	André	ANJOS	2022-09-01	2023-03-31
CN	CHI	Xuemin	Sylvain	CALINON	2022-11-01	2024-10-31
IT	CILLARI	Giacomo	Jérôme	KÄMPF	2022-02-28	2022-08-28
CH	COSTA	Alessandro	Sébastien	MARCEL	2022-02-21	2022-06-21
CH	DÉLITROZ	Maxime	Raphaëlle	LUISIER	2022-02-14	2022-08-15
CN	DONG	Yifei	Sylvain	CALINON	2022-09-15	2022-11-30
IR	FARKHONDEH	Arya	Jean-Marc	ODOBEZ	2022-10-01	2023-06-30
CH	FOURNIER	Lisa	Raphaëlle	LUISIER	2022-09-19	2023-01-23
ES	GARCIA GIRALDO	Juan	André	FREITAS	2022-06-13	2022-09-13
FR	GERVAISE	Lara	Jean-Marc	ODOBEZ	2021-11-01	2022-03-15
BE	GEVERS	Louis	Jean-Marc	ODOBEZ	2021-11-03	2022-03-31
CH	GINDROZ	Mickael	Sylvain	CALINON	2022-02-14	2022-06-17
CH	GIROUD	Mathieu	Jérôme	KÄMPF	2022-08-01	2022-08-31
CN	HUANG	Junduan	Sébastien	MARCEL	2022-12-15	2024-12-14
DK	IBSEN	Mathias	Sébastien	MARCEL	2022-01-10	2022-07-10
CH	JAUBERT	Côme	Raphaëlle	LUISIER	2022-02-21	2022-07-21
CN	JIANG	Xiaowen	Sylvain	CALINON	2022-09-01	2022-12-31
KR	KIM	Haeeun	Daniel	GATICA-PEREZ	2022-03-01	2022-07-31
CH	KOEHN	Edouard Erwan	Raphaëlle	LUISIER	2022-08-01	2023-01-31
FR	KUHN	Emilie	Sébastien	MARCEL	2022-02-21	2022-06-21
NO	LANG	Inga	Lonneke	VAN DER PLAS	2021-10-25	2022-04-25
CH	LOYE	Lena	Raphaëlle	LUISIER	2022-02-22	2022-06-04
MX	LUÉVANO GARCÍA	Luis Santiago	Sébastien	MARCEL	2022-09-01	2022-11-30
CH	MÄDER	Aurel	Daniel	GATICA-PEREZ	2021-10-01	2022-01-31
CH	MARGUET	Alexandre	Michael	LIEBLING	2022-02-21	2022-07-31
	MEADOWS	Jordan	André	FREITAS	2022-04-15	2022-08-15

CH	MONNET	Stephen	Sylvain	CALINON	2022-07-11	2022-09-16
CH	MORAIS	Antonio	André	ANJOS	2021-08-30	2022-02-25
MT	MUSCAT	Amanda	Petr	MOTLICEK	2022-09-01	2022-10-31
CH	PILLET	Maxime	Sylvain	CALINON	2022-02-25	2022-06-25
FR	PITON	Timothy	Mathew	MAGIMAI DOSS	2022-08-01	2022-09-30
FR	POCARD	Valentin	Philip	GARNER	2022-11-01	2023-04-30
CH	SALAMIN	Chloé	Petr	MOTLICEK	2022-06-27	2022-08-28
CH	SCHNEUWLY	Christelle	Raphaëlle	LUISIER	2022-08-01	2023-01-31
CH	STEL	Lucas	Olivier	BORNET	2022-09-15	2023-01-31
IR	TAVASSOLI	Mehrdad	Sylvain	CALINON	2022-05-15	2022-08-16
LK	THAYAPARAN	Mokanarangan	André	FREITAS	2021-10-01	2022-07-18
CN	TI	Boyang	Sylvain	CALINON	2021-02-24	2022-01-23
CH	TRUSCELLO	Léonard	Petr	MOTLICEK	2022-07-04	2022-10-19
IT	VALENTINO	Marco	André	FREITAS	2021-09-01	2022-05-14
CH	VAUCHER	Karine	Sébastien	MARCEL	2022-08-02	2023-01-31
IN	YADAV	Sarthak	Mathew	MAGIMAI DOSS	2022-04-01	2022-09-30

4.5 Visitors

Nationality	Last name	First name	Supervisor		Start	End (estimated)
IR	HOSSEINI KIVANANI	Nina	Ina	KODRASI	2022-07-11	2022-10-31
CA	ALI	Abid	Jean-Marc	ODOBEZ	2022-03-21	2022-04-01

5 Active and Granted Projects in 2022

An overview of the projects that have been active during the year 2022 is presented in Section 5.1. The projects are grouped in three categories, namely National Research Projects, European and International Research Projects and Industry-oriented Projects.

Section 5.2 presents the list of projects accepted during 2022 but starting in the following year.

5.1 Projects in Progress during 2022

5.1.1 National Research Projects

[1] Name **2000LAKES** (alpine research and citizen science toward the microbial conservation of high-mountain lakes in Switzerland)
Funding CLIMACT initiative
Coordinator EPFL
Duration 2022.03.01 - 2023.12.31
Partner(s) University of Lausanne; EPFL; Idiap Research Institute

[2] Name **AI4AUTISM2** (Digital Phenotyping of Autism Spectrum Disorders in Children)
Funding SNF - Sinergia
Coordinator University of Geneva
Duration 2021.11.01 - 2025.10.31
Partner(s) SUPSI; Idiap Research Institute

[3] Name **CAD4IED** (Computer assisted detection and grading of inflammatory eye diseases via fluorescein angiograms and novel biomarkers)
Funding Idiap Research Institute
Coordinator Idiap Research Institute
Duration 2022.10.01 - 2023.09.30
Partner(s) Fondation Asile des aveugles Hôpital Jules-Gonin; Lucerne Cantonal Hospital

[4] Name **CHASPEEPRO** (Characterisation of motor speech disorders and processes)
Funding SNF - Sinergia
Coordinator University of Geneva
Duration 2021.12.01 - 2025.11.30
Partner(s) Idiap Research Institute; HUG; New Sorbonne University Paris 3

[5] Name **C-LING** (Towards Creative systems with LINGuistic modelling)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2022.09.01 - 2026.08.31
Partner(s) -

[6] Name **CODIMAN** (A future that works: Cobotics, digital skills and the re-humanization of the workplace)
Funding SNF - NRP77
Coordinator Bern University of Applied Sciences
Duration 2020.05.01 - 2024.04.30
Partner(s) Idiap Research Institute

[7] Name **COMPBIO** (Computational biomicroscopy: advanced image processing methods to quantify live biological systems)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.04.01 - 2022.09.30
Partner(s) -

[8] Name **CORTI** (Computational Reduction for Training and Inference)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2020.03.01 - 2022.06.30
Partner(s) -

[9] Name **DOMAT** (On-demand Knowledge for Document-level Machine Translation)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.10.01 - 2022.09.30
Partner(s) HES-SO Master - Vaud

[10] Name **EMIL** (Emotion in the loop – a step towards a comprehensive closed-loop deep brain stimulation in Parkinson's disease)
Funding SNF - Bridge Discovery
Coordinator University of Bern
Duration 2021.05.01 - 2025.04.30
Partner(s) Idiap Research Institute; CSEM

[11] Name **EVOLANG** (Evolving Language)
Funding SNF - NCCR 5th series
Coordinator University of Zurich
Duration 2020.06.01 - 2024.05.31
Partner(s) University of Geneva; University of Neuchâtel; University of Basel; University of Lausanne; Idiap Research Institute; EPFL; Zurich University of the Arts; University of Fribourg; ETHZ

[12] Name **EYE-TRACKING** (Creating a Multilingual Eye-Tracking Corpus for Human and Machine-Based Language Processing)
Funding University of Zurich
Coordinator University of Zurich
Duration 2022.09.01 - 2025.08.31
Partner(s) Idiap Research Institute

[13] Name **HEAP** (Human-Guided Learning and Benchmarking of Robotic Heap Sorting)
Funding SNF - ERA-NET
Coordinator University of Lincoln
Duration 2019.05.01 - 2022.05.31
Partner(s) Italian Institute of Technology; Vienna University of Technology; INRIA; Idiap Research Institute

[14] Name **INTREPID** (Automated interpretation of political and economic policy documents: Machine learning using semantic and syntactic information)
Funding SNF - Sinergia
Coordinator Graduate Institute of International and Development Studies
Duration 2019.01.01 - 2023.12.31
Partner(s) Idiap Research Institute

[15] Name **IPEQ** (Uncertainty quantification and efficient design of experiments for data- and simulation-driven inverse problem solving)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.11.01 - 2022.10.31
Partner(s) University of California, Davis

[16] Name **LAOS** (Learning Representations of Abstraction in Text)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.10.01 - 2023.03.31
Partner(s) -

[17] Name **LEARN-REAL** (LEARNing physical manipulation skills with simulators using REAListic variations)
Funding SNF - ERA-NET
Coordinator Idiap Research Institute
Duration 2019.04.01 - 2023.03.31
Partner(s) Lyon Central School; Italian Institute of Technology

[18] Name **NAST** (Neural Architectures for Speech Technology)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2020.02.01 - 2024.01.31
Partner(s) -

[19] Name **NATAI** (The Nature of Artificial Intelligence)
Funding SNF - Agora
Coordinator Idiap Research Institute
Duration 2020.12.01 - 2023.11.30
Partner(s) Musee de la Main UNIL/CHUV

[20] Name **NEUMATH** (NeuMath: Neural Discourse Inference over Mathematical Texts)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2022.04.01 - 2025.03.31
Partner(s) -

[21] Name **NEUROCIRT** (Investigating the role of cytoplasmic intronic sequences in ALS pathogenesis)
Funding SNF - Division III
Coordinator Idiap Research Institute
Duration 2022.09.01 - 2026.08.31
Partner(s) -

[22] Name **NKBP** (Deep Learning Models for Continual Extraction of Knowledge from Text)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2020.10.01 - 2024.09.30
Partner(s) Catholic University of Leuven

[23] Name **ROSALIS** (Robot skills acquisition through active learning and social interaction strategies)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.04.01 - 2022.03.31
Partner(s) -

[24] Name **SAFER** (reSponsible fAir FacE Recognition)
Funding Hasler Responsible AI
Coordinator Idiap Research Institute
Duration 2022.03.01 - 2026.02.28
Partner(s) SICPA; University of Zurich

[25] Name **SHISSM** (Sparse and hierarchical Structures for Speech Modeling)
Funding SNF - Division II
Coordinator Idiap Research Institute
Duration 2018.03.01 - 2022.10.31
Partner(s) -

[26] Name **SMILE-II** (SMILE-II Scalable Multimodal sign language technology for sIgn language Learning and assessmEnt Phase-II)
Funding SNF - Sinergia
Coordinator Idiap Research Institute
Duration 2021.01.01 - 2024.12.31
Partner(s) University of Applied Sciences of Special Needs Education; University of Surrey; University of Zurich

[27]	Name	**STEADI** (Storytelling and first impressions in face-to-face and algorithm-powered digital interviews)
	Funding	SNF - Division I
	Coordinator	University of Neuchâtel
	Duration	2021.02.01 - 2025.01.31
	Partner(s)	Idiap Research Institute; University of Lausanne

[28]	Name	**SWITCH** (Learning by Switching Roles in Physical Human-Robot Collaboration)
	Funding	SNF - Division II
	Coordinator	Idiap Research Institute
	Duration	2021.03.01 - 2024.02.29
	Partner(s)	Jozef Stefan Institute

[29]	Name	**TIPS** (Towards Integrated processing of Physiological and Speech signals)
	Funding	SNF - Division II
	Coordinator	Idiap Research Institute
	Duration	2019.12.01 - 2023.11.30
	Partner(s)	CSEM; Coaching and Moderation

5.1.2 European and International Research Projects

[1] Name **AI4MEDIA** (A European Excellence Centre for Media, Society and Democracy)
Funding H2020
Coordinator Centre for Research and Technology Hellas
Duration 2020.09.01 - 2024.08.31
Partner(s) University of Trento; Queen Mary University of London; University of Amsterdam; HES-SO Valais; Catholic University of Leuven; Commissariat à l'Energie Atomique et aux Energies Alternatives; Joanneum Research; Fraunhofer; Idiap Research Institute; Globaz SA; Interdigital R&D France; F6S; Modl.ai; Imagga Technologies; Grassroots; Athens Technology Center; IBM Ireland ; Ircam; Côte d'Azur University; University of Malta; VRT; University of Bucarest; University of Florence; Aristotle University of Thessaloniki; Italian Radio Television; National Research Council of Italy; Deutsche Welle; Netherlands Institute for Sound and Vision

[2] Name **ATCO2** (Automatic collection and processing of voice data from air-traffic communications)
Funding H2020
Coordinator Idiap Research Institute
Duration 2019.11.01 - 2022.02.28
Partner(s) Brno University of Technology; Saarland University; Romagna Tech; Replay Well; ELDA; Open Sky Network

[3] Name **COLLABORATE** (Co-production CeLL performing Human-Robot Collaborative AssEmbly)
Funding H2020
Coordinator Aristotle University of Thessaloniki
Duration 2018.10.01 - 2022.03.31
Partner(s) University of Genoa; Pratt & Whitney Rzeszów; University of Patras; Blue Ocean Robotics APS; Jozef Stefan Institute; Arcelik; Catholic University of Leuven; Centre for Research and Technology Hellas; ARMINES; Fiat SCPA Research Centre; ASTI Mobile Robotics; Kolektor; Idiap Research Institute

[4] Name **CRITERIA** (Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks)
Funding H2020
Coordinator University of Hannover
Duration 2021.09.01 - 2024.08.31
Partner(s) General Inspectorate of Romanian Border Police; Swedish National Police Authority; University of Groningen; Hensoldt Analytics; Centre for Research and Technology Hellas; Malta Police Force; ARSIS Association for the Social Support of Youth; Idiap Research Institute; Knowledge and Innovation; webLyzard technology; European University of Cyprus; Estonian Police and Border Guard Board; University of Malta; Ministry of Interior Croatia

[5] Name **ENERMAPS** (Open Source Tools to Share, Compare, and Reuse Low-Carbon Energy Data)
Funding H2020
Coordinator CREM
Duration 2020.04.01 - 2022.06.30
Partner(s) Revolve Media; OpenAire; e-think; Idiap Research Institute; Vienna University of Technology; Eurac

[6] Name **GRAIL** (Generative Range and Altitude Identity Learning)
Funding USA IARPA
Coordinator University of Southern California
Duration 2021.11.12 - 2025.11.11
Partner(s) Idiap Research Institute

[7] Name **HAAWAII** (Highly Automated Air Traffic Controller Workstations with Artificial Intelligence Integration)
Funding H2020-SESAR
Coordinator DLR
Duration 2020.06.01 - 2022.11.30
Partner(s) NATS (En Route) ; Isavia; Austro Control ; Brno University of Technology; Idiap Research Institute; Croatian Air Control

[8] Name **ICARUS** (Innovative AppRoach to Urban Security)
Funding H2020
Coordinator Forum Europeen Pour La Securite Urbaine
Duration 2020.09.01 - 2024.08.31
Partner(s) Salzburg University of Applied Sciences ; Commune de Nice; Eurocircle Association; Makesense; Lisbon Municipal Police; University of Leeds; City of Torino; Globaz SA; Greek Center for Security Studies; City of Rotterdam; Riga Municipal Police; Landeshauptstadt Stuttgart; University of Salford; Panteion University of Social and Political Sciences; Erasmus University Rotterdam; Ethical and Legal Plus; Idiap Research Institute

[9] Name **INTELLIMAN** (AI-Powered Manipulation System for Advanced Robotic Service, Manufacturing and Prosthetics)
Funding Horizon Europe
Coordinator University of Bologna
Duration 2022.09.01 - 2026.02.28
Partner(s) University of Genoa; National Institute for Industrial Accidents Insurance Inail; Elvez; Bavarian Research Alliance Gmbh; University of Campania Luigi Vanvitelli; Eurecat Foundation; Idiap Research Institute; Technical University of Catalonia; DLR; University of Zurich; University of Erlangen-Nuremberg; Ocado Innovation Limited

[10] Name **MEMMO** (Memory of Motion)
Funding H2020
Coordinator CNRS
Duration 2018.01.01 - 2022.06.30
Partner(s) University of Edinburgh; Idiap Research Institute; University of Oxford; Max Planck Society; AIRBUS; Wandercraft; Pal Robotics S.L.; Centre de medecine physique et de readaptation; Costain Group PLC

[11] Name **ROXANNE** (Real time network, text, and speaker analytics for combating organized crime)
Funding H2020
Coordinator Idiap Research Institute
Duration 2019.09.01 - 2022.12.31
Partner(s) Trilateral; ADDITESS; Greek Center for Security Studies; Lithuanian Forensic Science Center; Phonexia; Brno University of Technology; University of Hannover; Police Service of Northern Ireland; INTERPOL; Hensoldt Analytics; Netherlands Forensic Institute; Saarland University; Police of the Czech Republic; Capgemini ; Ministry of Internal Affairs of Romania; Catholic University of the Sacred Heart; Aegis; Airbus Defence and Space; Ministry of Interior Croatia; ITML; Ministry of Public Security of Israel; Hellenic Police; National Police Force of Ireland; ZITIS

[12] Name **SESTOSENSO** (Physical Cognition for Intelligent Control and Safe Human-Robot Interaction)
Funding Horizon Europe
Coordinator University of Genoa
Duration 2022.10.01 - 2025.09.30
Partner(s) Fiat SCPA Research Centre; Free University of Bolzano; Ocado Innovation Limited; Inertia Technology B.V.; University of Latvia; Rise Research Institutes of Sweden AB; Institut Franco-Allemand de Recherches de Saint-Louis; University of Oxford; University of Ljubljana; University of Zaragoza; University of Bologna; Centre for Research and Technology Hellas; Idiap Research Institute

[13] Name **SOTERIA** (uSer-friendly digiTal sEcured peRsonal data and prIvacy plAtform)
Funding H2020
Coordinator Ariadnext
Duration 2021.10.01 - 2024.09.30
Partner(s) INRIA; Asociatia Infocons; Scytl Election Technologies; Stelar Security Technology Law Research UG; Noria Onlus; Erdyn Atlantique; Audencia Business School; Idiap Research Institute; ipcenter.at; Basque Health Service Osakidetza; Autonomous University of Barcelona; Catholic University of Leuven

[14] Name **TAPAS** (Training Network on Automatic Processing of PAthological Speech)
Funding H2020-MSCA
Coordinator Idiap Research Institute
Duration 2017.11.01 - 2022.09.30
Partner(s) Philips; University of Sheffield; Interuniversity Microelectronics Centre; University of Augsburg; Radboud University Nijmegen; The Netherlands Cancer Institute; Institute for Systems Engineering of Lisbon; University of Erlangen-Nuremberg ; Antwerpen University Hospital; Institut de Recherche en Informatique de Toulouse; Ludwig Maximilian University of Munich

[15] Name **TRESPASS-ETN** (TRaining in Secure and PrivAcy-preserving biometricS)
Funding H2020-MSCA
Coordinator EURECOM
Duration 2020.01.01 - 2023.12.31
Partner(s) University of Groningen; Idiap Research Institute; Darmstadt University of Applied Sciences; Chalmers University of Technology; Catholic University of Leuven; Autonomous University of Madrid

[16] Name **WENET** (The Internet of US)
Funding H2020
Coordinator University of Trento
Duration 2019.01.01 - 2023.06.30
Partner(s) Potosino Institute of Scientific and Technological Research; London School of Economics and Political Science; Eberhard Karls University of Tübingen; Catholic University of Asunción; Martel; Ben Gurion University of the Negev; Open University of Cyprus; Idiap Research Institute; U-Hopper; University of Aalborg; Jilin University; Amrita University; National University of Mongolia

5.1.3 Industry-oriented Projects

[1] Name **ABROAD** (Development of an NLP tool for selecting the potential sources of novel antibiotic active against multiresistant microbes)
Funding The Ark: NPR
Coordinator CimArk
Duration 2022.09.01 - 2024.02.29
Partner(s) Idiap Research Institute; IHMA Europe Sàrl - BioArk; Inflamalps SA

[2] Name **AI_SENSOR** (Sensor Fusion and Active Sensing for World-View Understanding)
Funding Industrial
Coordinator Idiap Research Institute
Duration 2020.02.03 - 2024.01.31
Partner(s) ams International

[3] Name **CANDY** (ContActless finger veiN recognition and presentation attack Detection on-the-flY)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.03.01 - 2022.11.30
Partner(s) Global ID SA

[4] Name **CMM** (Conversation Member Match)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.07.01 - 2022.06.30
Partner(s) Speak & Lunch

[5] Name **EGUZKI** (Eguzki : Programme de simulation de réseaux de chauffage à distance basé sur l'intelligence artificielle pour la résolution rapide et prédictive de réseaux complexes bouclés.)
Funding SFOE
Coordinator RWB Valais SA
Duration 2020.09.01 - 2023.08.30
Partner(s) Idiap Research Institute; Satom SA; Altis Groupe SA; Oiken SA

[6] Name **EPARTNERS4ALL** (a (personalized and) blended care solution with virtual buddy for child health)
Funding Eureka (Innosuisse)
Coordinator Netherlands Organisation for Applied Scientific Research
Duration 2021.11.15 - 2024.05.15
Partner(s) Therapieland; MedVision 360 BV; Organization for Fire Services, Crisis Management and Public Health; Eyeware; Leiden University Medical Center; Topicus Healthcare B.V.; Delft University of Technology; Interactive Robotics B.V.; Bern University of Applied Sciences; Idiap Research Institute; Xpert Health Cory Care

[7] Name **GAZESENSESCREEN** (GazeSense Screen)
Funding The Ark
Coordinator Eyeware
Duration 2020.03.01 - 2022.05.31
Partner(s) Idiap Research Institute

[8] Name **HARDENING** (Heterogeneous face recognition for unified identity management)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2021.02.01 - 2023.04.30
Partner(s) Facedapter Sàrl

[9] Name **IICT** (Inclusive Information and Communication Technologies)
Funding Innosuisse (Flagship)
Coordinator University of Zurich
Duration 2022.03.14 - 2026.03.13
Partner(s) Federal Office for the Equality of Persons with Disabilities; Zurich Insurance; Federal Office for Civil Protection; Swiss Federation of the Deaf; CFS GmbH Capito; Swiss Txt; University of Applied Sciences of Special Needs Education; Institut Icare; Idiap Research Institute

[10] Name **IVECT** (Impact of greening on the energy balance and thermal comfort of buildings and districts)
Funding SFOE
Coordinator HES-SO Valais
Duration 2020.12.01 - 2023.11.30
Partner(s) Idiap Research Institute; CREM; Etat du Valais; City of Zurich; Zurich University of Applied Sciences

[11] Name **LUCIDELES** (Leveraging User-Centric Intelligent Daylight and Electric Lighting for Energy Saving)
Funding SFOE
Coordinator University of Fribourg
Duration 2020.02.01 - 2023.12.31
Partner(s) Regent ; Idiap Research Institute

[12] Name **MALAT** (Machine Learning for Air Traffic)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.03.01 - 2023.02.28
Partner(s) SkySoft-ATM

[13] Name **MARGIN** (Multi-modal federated age verification)
Funding Innosuisse
Coordinator Idiap Research Institute
Duration 2020.07.01 - 2022.07.30
Partner(s) Privately SA; EPFL

[14] Name **P3** (Press Pressure Prediction)
Funding Innosuisse
Coordinator HES-SO Valais
Duration 2021.06.01 - 2022.11.30
Partner(s) Idiap Research Institute; Constellium Valais SA

[15] Name **STARFISH** (Safety and Speech Recognition with Artificial Intelligence in the Use of Air Traffic Control)
Funding Industrial
Coordinator Idiap Research Institute
Duration 2020.10.31 - 2022.09.30
Partner(s) Deutsches Zentrum für Luft und Raumfahrt

[16] Name **WAVE2-96** (H2020-SESAR-PJ.10-W2-Solution 96)
Funding Industrial
Coordinator Idiap Research Institute
Duration 2020.10.01 - 2023.01.31
Partner(s) Deutsches Zentrum für Luft und Raumfahrt

5.2 Projects Awarded in 2022 and Starting in 2023

[1] Name **MELAS** (Alternative Splicing and Polyadenylation from single-cell RNA sequencing towards tumor subpopulation identification in melanoma)
Funding Novartis Foundation
Coordinator Idiap Research Institute
Duration 2023.05.01 - 2024.04.30
Partner(s) Novartis Research Foundation

[2] Name **MELAS-KPI** (Towards the development of predictive biomarkers for patient stratification and immunotherapy response in cancer)
Funding Idiap Research Institute
Coordinator Idiap Research Institute
Duration 2023.04.01 - 2024.03.31
Partner(s) CHUV; Novartis Institutes for BioMedical Research; Hôpital de Sion

[3] Name **MULTIPLEYE** (Enabling multilingual eye-tracking data collection for human and machine language processing research)
Funding COST - H2020
Coordinator University of Copenhagen
Duration 2023.01.01 - 2025.12.31
Partner(s) Idiap Research Institute

[4] Name **PREP-VIDAL-SNSF** (A personalized speech recognition framework for audio messaging on the edge)
Funding SNSF
Coordinator Idiap Research Institute
Duration 2023.05.01 - 2024.04.30
Partner(s) -

[5] Name **TRACY** (A big-data analyTics from base-stations Registrations And Cdrs e-evidence sYstem)
Funding Horizon Europe
Coordinator PERFORMANCE
Duration 2023.04.01 - 2025.03.31
Partner(s) Hellenic Police; University of Applied Sciences for Public Administration and Legal Affairs in Bavaria; COSMOTE; Idiap Research Institute; POLICE_ROMANIA_GL; POLICE_MOLDOVA; Greek Center for Security Studies; TIMELEX

[6] Name **UNDERSPEC-ROBUSTNESS** (Addressing Underspecification for Improved Fairness and Robustness in Conversational AI)
Funding Amazon Research Awards
Coordinator Idiap Research Institute
Duration 2023.06.01 - 2024.05.31
Partner(s) -

[7] Name **UNIDIVE** (Universality, diversity and idiosyncrasy in language technology)
Funding COST - H2020
Coordinator Université Paris-Saclay
Duration 2023.01.01 - 2025.12.31
Partner(s) Idiap Research Institute

6 List of Publications in 2022

6.1 Book Chapters

[1] M. Á. Álvarez-Carmona, E. VILLATORO-TELLO, L. V. Pineda, and M. Montes-y Gómez, "Classifying the social media author profile through a multimodal representation," in *Intelligent Technologies: Concepts, Applications, and Future Directions. Studies in Computational Intelligence*, ser. 7092, vol. 1028, Springer, May 2022.

[2] N. Baker, R. Belmonte Monteiro, A. Boccalatte, K. Bouty, J. Brozovsky, C. Caliot, R. Campamà Pizarro, R. Compagnon, A. Czachura, G. Desthieux, M. Formolli, S. Giroux-Julien, V. Guillot, G. Benjamin, C. Hachem-Vermette, E. Herman, O. Alarcon Herrera, J. Kämpf, G. Lobaccaro, C. Ménézo, M. Musy, G. Peronato, A. J. Petersen, A. Rodler, K. Singh, V. Sjöberg, M. Snow, J. Tjetland, and Y. Wang, "Identification of existing tools and workflows for solar neighborhood planning," in *SHC Task 63: Solar Neighborhood Planning, Subtask C: Solar Planning Tools*, J. Kanters and M. Thebault, Eds., IEA, 2022.

[3] S. Burdisso, L. Cagnina, M. Errecalde, and M. Montes-y Gómez, "Two simple and domain-independent approaches for early detection of anorexia," in *Early Detection of Mental Health Disorders by Social Media Monitoring: The First Five Years of the eRisk Project*, 202nd ed., Springer International Publishing, 2022, pp. 159–182.

[4] S. Burdisso, Z.-G. Juan, E. VILLATORO-TELLO, M. Fajcik, M. Singh, P. Smrz, and P. Motlicek, "Idiapers @ causal news corpus 2022: Efficient causal relation identification through a prompt-based few-shot approach," in *The 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE @ EMNLP 2022)*, ACL Anthology (https://aclanthology.org/2022.case-1.9), Dec. 2022.

[5] S. R. Dash, S. Parida, E. VILLATORO-TELLO, B. Acharya, and O. Bojar, *Natural Language Processing in Healthcare*, 1st, S. R. Dash, S. Parida, E. VILLATORO-TELLO, B. Acharya, and O. Bojar, Eds. Taylor and Francis Groups, Mar. 2022, The list of authors is too long, hence only the main editors are named as authors.

[6] L. E. Medina Rios, S. Ruiz-Correa, D. Santani, and D. Gatica-Perez, "Who sees what? examining urban impressions in global south cities," in *Human Perception of Visual Information: Psychological and Computational Perspectives*, Springer, 2022.

[7] E. VILLATORO-TELLO, S. Madikeri, P. Motlicek, A. Ganapathiraju, and A. V. Ivanov, "Expanded lattice embeddings for spoken document retrieval on informal meetings," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Jul. 2022.

6.2 Articles in Scientific Journals

[1] M. Axelsson, R. Oliveira, M. Racca, and V. Kyrki, "Social robot co-design canvases: A participatory design framework," *ACM Transactions on Human-Robot Interaction*, vol. 11, no. 1, 2022.

[2] C. Basurto, M. Papinutto, M. Colombo, R. Boghetti, K. Reutter, J. Nembrini, and J. Kämpf, "Integrating daylight with general and task lighting: A longitudinal in-the-wild study in individual and open space working areas," *Solar Energy Advances*, vol. 2, Nov. 2022.

[3] A. Bittar and P. N. Garner, "A surrogate gradient spiking baseline for speech command recognition," *Frontiers in Neuroscience*, Aug. 2022.

[4] H. Bourlard and S. H. Kabil, "Autoencoders reloaded," *Springer Biological Cybernetics*, Jun. 2022.

[5] H. Burke, A. Freeman, P. O'Reagan, O. Wysocki, A. Freitas, and e. al., "Biomarker identification using dynamic time warping analysis: A longitudinal cohort study of covid-19 patients in a uk tertiary hospital," *BMJ Open*, 2022.

[6] Z. Dong, Z. Li, Y. Yan, S. Calinon, and F. Chen, "Passive bimanual skills learning from demonstration with motion graph attention networks," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 2, pp. 4917–4923, 2022.

[7] S. P. Dubagunta, R. J. J. H. van Son, and M. Magimai.-Doss, "Adjustable deterministic pseudonymization of speech," *Computer, Speech, and Language*, vol. 72, Mar. 2022, Open Access.

[8] C. Fougeron, I. Kodrasi, and M. Laganaro, "Differentiation of motor speech disorders through the seven deviance scores from monpage-2.0.s," *Brain Sciences*, vol. 12, no. 11, pp. 1471–1487, Oct. 2022.

[9] A. Freeman, A. Watson, P. O'Reagan, O. Wysocki, H. Burke, A. Freitas, and e. al., "Wave comparisons of clinical characteristics and outcomes of covid-19 admissions - exploring the impact of treatment and strain dynamics," *Journal of Clinical Virology*, Jan. 2022.

[10] L. Fritz, U. Vilsmaier, G. Clement, L. Daffe, A. Pagani, M. Pang, D. Gatica-Perez, V. Kaufmann, M. S. Delefosse, and C. R. Binder, "Explore, engage, empower: Methodological insights into a transformative mixed methods study tackling the covid-19 lockdown.," *Humanities and Social Sciences Communications*, May 2022.

[11] H. Frost, D. Graham, L. Carter, P. O'Reagan, D. Landers, and A. Freitas, "Patient attrition in molecular tumour boards: A systematic review," *British Journal of Cancer*, 2022.

[12] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, and I. Ben Ayed, "State-of-the-art retinal vessel segmentation with minimalistic models," *Nature Scientific Reports*, vol. 12, no. 6174, Apr. 2022.

[13] T. Gentilhomme, M. Villamizar, J. Corre, and J.-M. Odobez, "Towards smart pruning: Vinet, a deep-learning approach for grapevine structure estimation," *Computers and Electronics in Agriculture*, 2022, Submitted.

[14] A. George, A. Mohammadi, and S. Marcel, "Prepended domain transformer: Heterogeneous face recognition without bells and whistles," *IEEE Transactions on Information Forensics and Security*, 2022.

[15] J. Henderson and F. Fehr, "A variational autoencoder for transformers with nonparametric variational information bottleneck," *Arxiv*, 2022.

[16] A. Hume, L. Cernuzzi, J. L. Zarza, I. Bison, and D. Gatica-Perez, "Analysis of the big-five personality traits in the chatbot "uc - paraguay"," *CLEI electronic journal*, vol. 25, no. 2, May 2022.

[17] J. Jankowski, M. Racca, and S. Calinon, "From key positions to optimal basis functions for probabilistic adaptive control," *IEEE Robotics and Automation Letters*, Jan. 2022.

[18] K. Kotwal, S. Bhattacharjee, P. Abbet, Z. Mostaani, H. Wei, X. Wenkang, Z. Yaxi, and S. Marcel, "Domain-specific adaptation of cnn for detecting face presentation attacks in nir," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2022.

[19] V. Krivokuca and S. Marcel, "Biometric template protection for neural-network-based face recognition systems: A survey of methods and evaluation techniques," *ArXiv*, 2022, Version 2 – After addition of BTP vs B-PET explanation + 2 new references.

[20] R. Lee, O. Wysocki, A. Freitas, and e. al., "Establishment of coronet, covid-19 risk in oncology evaluation tool, to identify cancer patients at low versus high risk of severe complications of covid-19 infection upon presentation to hospital," *Clinical Cancer Informatics*, 2022.

[21] J. Liu, Y. Chen, Z. Dong, S. Wang, S. Calinon, M. Li, and F. Chen, "Robot cooking with stir-fry: Bimanual non-prehensile manipulation of semi-fluid objects," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 2, pp. 5159–5166, 2022.

[22] y. liu, J. Dong, T.-A. Pham, F. Marelli, and M. Unser, "Mechanical artifacts in optical projection tomography: Classification and automatic calibration," *Opt. Continuum*, vol. 1, no. 12, pp. 2577–2589, Dec. 2022.

[23] T. Löw, J. Maceiras, and S. Calinon, "Drozbot: Using ergodic control to draw portraits," *IEEE Robotics and Automation Letters*, p. 7, 2022.

[24] R. Luisier, C. Andreassi, and A. Riccio, "The rna binding proteome of axonal mrnas in sympathetic neurons," *BioRxiv*, 2022.

[25] F. Mai, A. Pannatier, F. Fehr, H. Chen, F. Marelli, F. Fleuret, and J. Henderson, "Hypermixer: An mlp-based green ai alternative to transformers," *Arxiv*, 2022.

[26] L. B. Meegahapola, W. Bangamuarachchi, A. Chamantha, S. Ruiz-Correa, I. Perera, and D. Gatica-Perez, "Sensing eating events in context: A smartphone-only approach," *IEEE Access*, vol. 10, May 2022.

[27] L. B. Meegahapola, W. Droz, A. de Götzen, P. KUN, C. Nuttakki, S. Diwakar, S. Ruiz-Correa, D. Song, H. Xu, M. Bidoglia, G. Gaskell, A. Chagnaa, A. Ganbold, T. Zundui, C. Caprini, D. Miorandi, A. Hume, J. L. Zarza, L. Cernuzzi, I. Bison, M. R. Britez, M. Busso, R. Chenu-Abente, C. Gunel, F. Giunchiglia, L. Schelenz, and D. Gatica-Perez, "Generalization and personalization of mobile sensing-based mood inference models: An analysis of college students in eight countries," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 6, no. 4, Dec. 2022.

[28] P. O'Reagan, A. Freitas, and e. al., "Digital ecmt cancer trial matching tool, an open source research application to support oncologists in the identification of precision medicine clinical trials," *JCO Clinical Cancer Informatics*, 2022.

[29] M. Papinutto, R. Boghetti, M. Colombo, C. Basurto, K. Reutter, D. Lalanne, J. Kämpf, and J. Nembrini, "Saving energy by maximising daylight and minimising the impact on occupants: An automatic lighting system approach," *Energy and Buildings*, 2022.

[30] M. Pernon, F. Assal, I. Kodrasi, and M. Laganaro, "Perceptual classification of motor speech disorders: The role of severity, speech task, and listener's expertise," *Journal of Speech, Language, and Hearing Research*, 2022.

[31] M. Petric-Howe, H. Crerar, J. Neeves, J. Harley, G. E. Tyzack, P. Klein, A. Ramos, R. Patani, and R. Luisier, "Physiological intron retaining transcripts in the cytoplasm abound during human motor neurogenesis," *Genome Research*, 2022.

[32] T.-T. Phan, C. Michoud, L. Volpato, M. del Río Carral, and D. Gatica-Perez, "Health talk: Understanding practices of popular professional youtubers," *Int. Conf. on Mobile and Ubiquitous Multimedia*, Nov. 2022.

[33] E. Pignat, J. Silverio, and S. Calinon, "Learning from demonstration using products of experts: Applications to manipulation and task prioritization," *International Journal of Robotics Research*, vol. 41, no. 2, pp. 163–188, 2022.

[34] C. Roberts, J. M. E. Herzing, M. A. Manjon, P. Abbet, and D. Gatica-Perez, "Response burden and dropout in a probability-based online panel study – a comparison between an app and browser-based design," *Journal of Official Statistics*, C. Roberts, Ed., Dec. 2022.

[35] C. Roberts, J. M. E. Herzing, J. Sobrino Piazza, P. Abbet, and D. Gatica-Perez, "Data privacy concerns as a source of resistance to complete mobile data collection tasks via a smartphone app," *Journal of Survey Statistics and Methodology*, Jun. 2022.

[36] B. Schnell and P. N. Garner, "Investigating a neural all pass warp in modern tts applications," *Speech Communication*, vol. 138, pp. 26–37, Mar. 2022, Open Access.

[37] N. Schütz, S. E. J. Knobel, M. Single, B. Pais, V. Santschi, D. Gatica-Perez, P. Buluschek, P. Urwyler, S. M. Gerber, R. M. Müri, U. P. Mosimann, H. Saner, and T. Nef, "A systems approach towards remote health-monitoring in older adults: Introducing a zero-interaction digital exhaust," *Npj Digital Medicine*, vol. 5, no. Article 116, Aug. 2022.

[38] M. Shamsi, A. Larcher, L. Barrault, S. Meignier, Y. Prokopalo, M. Tahon, A. Mehrish, S. Petitrenaud, O. Galibert, S. Gaist, A. Anjos, S. Marcel, and M. Costa-Jussà, "Towards lifelong human assisted speaker diarization," *Computer, Speech, and Language*, Jul. 2022.

[39] S. Shetty, J. Silverio, and S. Calinon, "Ergodic exploration using tensor train: Applications in insertion tasks," *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 906–921, Mar. 2022.

[40] R. Siegfried and J.-M. Odobez, "Robust unsupervised gaze calibration using conversation and manipulation attention priors," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 1, p. 26, Jan. 2022.

[41] M. Thayaparan, M. Valentino, D. Mendes, J. Rozanova, and A. Freitas, "Diff-explainer: Differentiable convex optimization for explainable multi-hop inference," *Transactions of the Association for Computational Linguistics*, 2022.

[42] O. Wysocki, J. K. Davies, M. Vigo, A. C. Armstrong, D. Landers, R. Lee, and A. Freitas, "Assessing the communication gap between ai models and healthcare professionals: Explainability, utility and trust in ai-driven clinical decision-making," *Artificial Intelligence*, 2022.

[43] O. Wysocki, Z. Zhou, P. O'Reagan, D. Mendes, D. Landers, and A. Freitas, "Transformers and the representation of biomedical background knowledge," *Computational Linguistics*, 2022.

[44] O. Ziff, J. Neeves, J. Mitchell, G. E. Tyzack, C. M. Ruiz, N. McGranahan, R. Luisier, A. Chakrabarti, S. Boulton, G. Kelly, J. Humphrey, and R. Patani, "Meta-analysis of the amyotrophic lateral sclerosis spectrum uncovers genome instability," *BioRxiv*, 2022.

6.3 PhD Theses

[1] A. Costa, “Using synthetic fingerprint images to test the performance of an afis system,” PhD thesis, Université de Lausanne, Jul. 2022.

[2] T. S. Lembono, “Memory of motion for initializing optimization in robotics,” PhD thesis, École Polytechnique Fédérale de Lausanne, Jul. 2022.

[3] B. Schnell, “Controllability and interpretability in affective speech synthesis,” PhD thesis, EPFL, Feb. 2022.

[4] A. Vyas, “Efficient transformer-based speech recognition,” PhD thesis, École polytechnique fédérale de Lausanne, Oct. 2022.

6.4 Articles in Conference Proceedings

[1] A. Agrawal, D. Teney, and A. Nematzadeh, "Vision-language pretraining: Current trends and the future," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, 2022.

[2] D. Alonso del Barrio and D. Gatica-Perez, "How did europe's press cover covid-19 vaccination news? a five-country analysis," in *MAD '22: Proceedings of the 1st International Workshop on Multimedia AI against Disinformation*, Jun. 2022.

[3] C. Basurto, M. Papinutto, M. Colombo, R. Boghetti, K. Reutter, J. Nembrini, and J. Kämpf, "An exploratory interplay between daylight, general and task lighting for visual comfort and electricity savings in a personal office space," in *Proceedings of ISES and IEA SHC International Conference on Solar Energy for Buildings and Industry*, Kassel, Germany, Sep. 2022.

[4] A. Bhatnagar, N. Bhavsar, M. Singh, and P. Motlicek, "An end-to-end multilingual system for automatic minuting of multi-party dialogues," in *Pacific Asia Conference on Language, Information and Computation (PACLIC 36) , In proceedings of ACL Anthology*, 2022.

[5] A. Bhatnagar, N. Bhavsar, M. Singh, and P. Motlicek, "Deepcon: An end-to-end multilingual toolkit for automatic minuting of multi-party dialogues," in *Special Interest Group on Discourse and Dialogue (SIGDIAL 2022)*, 2022.

[6] A. Bhatnagar, N. Bhavsar, M. Singh, and P. Motlicek, "Hierarchical multi-task learning framework for isometric-speech language translation," in *ACL*, 2022.

[7] N. Bhavsar, A. Bhatnagar, M. Singh, and P. Motlicek, "Hmist: Hierarchical multilingual isometric speech translation using multi-task learning framework for automatic dubbing," in *PACIFIC ASIA CONFERENCE ON LANGUAGE, INFORMATION AND COMPUTATION (PACLIC 2022), In Proceedings of ACL Anthology*, 2022.

[8] N. Bhavsar, A. Bhatnagar, M. Singh, P. Motlicek, and T. Ghosal, "Team innovators at semeval-2022 for task 8: Multi-task training with hyperpartisan and semantic relation for multi-lingual news article similarity," in *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), NAACL 2022*, 2022.

[9] A. Bittar and P. N. Garner, "Bayesian recurrent units and the forward backward algorithm," in *Proc. Interspeech 2022*, Sep. 2022, pp. 4137–4141.

[10] A. Bogatu, N. Paton, M. Douthwaite, and A. Freitas, "Voyager: Data discovery for onboarding in data science," in *37th IEEE International Conference on Data Engineering (ICDE)*, 2022.

[11] E. Bouton-Bessac, D. Gatica-Perez, and L. B. Meegahapola, "Your day in your pocket: Complex activity recognition from smartphone accelerometers," in *EAI Pervasive Health*, 2022.

[12] S. Burdisso, Z.-G. Juan, E. VILLATORO-TELLO, M. Fajcik, M. Singh, P. Smrz, and P. Motlicek, "Idiapers @ causal news corpus 2022: Efficient causal relation identification through a prompt-based few-shot approach," in *The 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE @ EMNLP 2022)*, Dec. 2022.

[13] C. Callejas-Hernández, E. Rivadeneira-Pérez, F. Sánchez-Vega, A. P. López-Monroy, and E. VILLATORO-TELLO, "The winning approach for the recommendation systems shared task @restmex 2022," in *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*, vol. 3202, Sep. 2022.

[14] G. Cillari, F. Fantozzi, A. Franco, and J. Kämpf, "Towards energy hubs: An innovative geographic information system based approach for cluster definition," in *ICREC 2022 Conference Proceedings*, E. Reports, Ed., 2022.

[15] L. Colbois and S. Marcel, "On the detection of morphing attacks generated by gans," in *21st International Conference of the Biometrics Special Interest Group (BIOSIG 2022)*, Sep. 2022.

[16] E. Courdier and F. Fleuret, “Borrowing from yourself: Faster future video segmentation with partial channel update,” in *International Conference on Pattern Recognition*, 2022.

[17] E. Courdier, P. T. Sivaprasad, and F. Fleuret, “Paumer: Patch pausing transformer for semantic segmentation,” in *33th British Machine Vision Conference 2022, London, UK, 21 - 24 November 2022*, 2022.

[18] S. P. Dubagunta, E. Moneta, E. Theocharopoulos, and M. Magimai.-Doss, “Towards automatic prediction of non-expert perceived speech fluency ratings,” in *ACM International Conference on Multimodal Interaction (ICMI Companion)*, Nov. 2022.

[19] M. Fajcik, M. Singh, Z.-G. Juan, E. VILLATORO-TELLO, S. Burdisso, P. Motlicek, and P. Smrz, “Idiapers @ causal news corpus 2022: Extracting cause-effect-signal triplets via pre-trained autoregressive language model,” in *The 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE @ EMNLP 2022)*, Dec. 2022.

[20] C. Fougeron, N. Audibert, I. Kodrasi, P. Janbakhshi, M. Pernon, N. Leveque, S. Borel, M. Laganaro, H. Bourlard, and F. Assal, “Comparison of 5 methods for the evaluation of intelligibility in mild to moderate french dysarthric speech,” in *Annual Conference of the International Speech Communication Association*, Sep. 2022.

[21] L. Coppieters de Gibson and P. N. Garner, “Low-level physiological implications of end-to-end learning for speech recognition,” in *Proc. Interspeech 2022*, 2022, pp. 749–753.

[22] H. Girgin, J. Jankowski, and S. Calinon, “Reactive anticipatory robot skills with memory,” in *The International Symposium on Robotics Research*, 2022.

[23] A. Gupta, S. Tafasca, and J.-M. Odobez, “A modular multimodal architecture for gaze target prediction: Application to privacy-sensitive settings,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022.

[24] H. Helmke, K. Ondřej, S. Shetty, H. Arilíusson, T. S. Simiganoschi, M. Kleinert, O. Ohneiser, h. Ehr, Z.-G. Juan, and P. Smrz, “Readback error detection by automatic speech recognition and understanding – results of haawaii project for isavia’s enroute airspace,” in *11th SESAR Innovation Days*, ser. 1, SESAR, vol. 1, Dec. 2022, p. 9.

[25] I. Hwang, S. Lee, Y. Kwak, S. J. Oh, D. Teney, J.-H. Kim, and B.-T. Zhang, “Selecmix: Debiased learning by contradicting-pair sampling,” in *Advances in Neural Information Processing Systems*, 2022.

[26] I. Hwang, S. Lee, Y. Kwak, S. J. Oh, D. Teney, J.-H. Kim, and B.-T. Zhang, “Selecmix: Debiased learning by mixing up contradicting pairs,” in *ICML Workshop on Spurious Correlations, Invariance and Stability*, 2022.

[27] N. Iuliia, Z.-G. Juan, A. Prasad, S. S. Sarfjoo, and P. Motlicek, “A two-step approach to leverage contextual data: Speech recognition in air-traffic communications,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2022.

[28] P. Janbakhshi and I. Kodrasi, “Adversarial-free speaker identity-invariant representation learning for automatic dysarthric speech classification,” in *Annual Conference of the International Speech Communication Association*, Sep. 2022.

[29] P. Janbakhshi and I. Kodrasi, “Experimental investigation on stft phase representations for deep learning-based dysarthric speech detection,” in *International Conference on Acoustics, Speech, and Signal Processing*, 2022.

[30] M. M. Johari, Y. Lepoittevin, and F. Fleuret, “Geonerf: Generalizing nerf with geometry priors,” in *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[31] S. H. Kabil and H. Bourlard, “From undercomplete to sparse overcomplete autoencoders to improve lf-mmi speech recognition,” in *Proceedings of Interspeech Conference*, 2022.

[32] D. Khanna, A. Bhatnagar, N. Bhavsar, M. Singh, and P. Motlicek, “Bio-medical multi-label scientific literature classification using lwan and dual-attention module,” in *PACIFIC ASIA CONFERENCE ON LANGUAGE, INFORMATION AND COMPUTATION (PACLIC 2022), In Proceedings of ACL Anthology*, 2022.

[33] D. Khanna, M. Singh, and P. Motlicek, “Idiaptiet@lt-edi-acl2022 : Hope speech detection in social media using contextualized bert with attention mechanism,” in *ACL*, 2022.

[34] P. Korshunov, A. Jain, and S. Marcel, “Custom attribution loss for improving generalization and interpretability of deepfake detection,” in *International Conference on Acoustics, Speech, and Signal Processing*, May 2022.

[35] P. Korshunov and S. Marcel, “Face anthropometry aware audio-visual age verification,” in *ACM Multimedia*, Oct. 2022.

[36] K. Kotwal and S. Marcel, “Fairness index measures to evaluate bias in biometric recognition,” in *International Conference on Pattern Recognition Workshops*, Aug. 2022.

[37] K. Kotwal and S. Marcel, “Residual feature pyramid network for enhancement of vascular patterns,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2022.

[38] I. Lang, L. van der Plas, M. Nissim, and A. Gatt, “Visually grounded interpretation of noun-noun compounds in english,” in *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, Association for Computational Linguistics, 2022.

[39] J. Linke, P. N. Garner, G. Kubin, and B. Schuppler, “Conversational speech recognition needs data? experiments with austrian german,” in *Proceedings of the 13th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, Jun. 2022, pp. 4684–4691.

[40] J. M. Loyola, H. Thompson, S. Burdisso, and M. Errecalde, “Unsl at erisk 2022: Decision policies with history for early classification,” in *CEUR Workshop Proceedings*, 2022.

[41] rabeeh karimi mahabadi, L. Zettlemoyer, J. Henderson, s. marzieh, lambert mathias, V. Stoyanov, and M. Yazdani, “Perfect: Prompt-free and efficient few-shot learning with language models,” in *ACL*, 2022.

[42] F. Mai and J. Henderson, “Bag-of-vectors autoencoders for unsupervised conditional text generation,” in *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Online, 2022, 468–488.

[43] E. Manino, J. Rozanova, D. Carvalho, A. Freitas, and L. Cordeiro, “Systematicity, compositionality and transitivity of deep nlp models: A metamorphic testing perspective,” in *Findings of the ACL*, 2022.

[44] A. Marfurt and J. Henderson, “Unsupervised token-level hallucination detection from summary generation by-products,” in *Workshop on Generation, Evaluation and Metrics (GEM)*, 2022.

[45] A. Marfurt, A. Thornton, D. Sylvan, L. van der Plas, and J. Henderson, “A corpus and evaluation for predicting semi-structured human annotations,” in *Workshop on Generation, Evaluation and Metrics (GEM)*, 2022.

[46] G. Marshall, C. Jay, and A. Freitas, “Why scholars are diagramming neural network models,” in *13th International Conference on the Theory and Application of Diagrams*, 2022.

[47] D. Mendes, M. Thayaparan, M. Valentino, J. Rozanova, and A. Freitas, “To be or not to be an integer? encoding variables for mathematical text,” in *Findings of the ACL*, 2022.

[48] G. Mercatali, V. Garg, and A. Freitas, “Symmetry-induced disentanglement on graphs,” in *Advances in Neural Information Processing Systems 35*, 2022.

[49] K. Micallef, A. Gatt, M. Tanti, L. van der Plas, and C. Borg, “Pre-training data quality and quantity for a low-resource language: New corpus and bert models for maltese,” in *Proceedings of the workshop on Deep Learning for Low-Resource NLP*, 2022.

[50] L. Miculicich and J. Henderson, “Graph refinement for coreference resolution,” in *Findings of Association for >Computational Linguistics: ACL 2022*, 2022.

[51] A. Mohammadshahi, V. Nikoulina, A. Berard, C. Brun, J. Henderson, and L. Besacier, “Small-100: Introducing shallow multilingual machine translation model for low-resource languages,” in *The 2022 Conference on Empirical Methods in Natural Language Processing*, 2022.

[52] A. Mohammadshahi, V. Nikoulina, A. Berard, C. Brun, J. Henderson, and L. Besacier, “What do compressed multilingual machine translation models forget?” In *The 2022 Conference on Empirical Methods in Natural Language Processing*, 2022.

[53] Z. Mostaani and M. Magimai.-Doss, “On breathing pattern information in synthetic speech,” in *Proceedings of Interspeech*, 2022.

[54] Z. Mostaani, R. Prasad, B. Vlasenko, and M. Magimai.-Doss, “Modeling of pre-trained neural network embeddings learned from raw waveform for covid-19 infection detection,” in *Proceedings of ICASSP*, 2022.

[55] C. Niklaus, A. Freitas, and S. Handschuh, “Shallow discourse parsing for open information extraction and text simplification,” in *3rd Workshop on Computational Approaches to Discourse (CODI) @ COLING*, 2022.

[56] D. Osorio-Roig, C. Rathgeb, H. Otroshi Shahreza, C. Busch, and S. Marcel, “Indexing protected deep face templates by frequent binary patterns,” in *Proceedings of the 2022 International Joint Conference on Biometrics (IJCB)*, Abu Dhabi, United Arab Emirates (UAE): IEEE, 2022.

[57] H. Otroshi Shahreza, V. Krivokuca, and S. Marcel, “Face reconstruction from deep facial embeddings using a convolutional neural network,” in *Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France: IEEE, 2022.

[58] H. Otroshi Shahreza, C. Rathgeb, D. Osorio-Roig, V. Krivokuca, S. Marcel, and C. Busch, “Hybrid protection of biometric templates by combining homomorphic encryption and cancelable biometrics,” in *Proceedings of the 2022 International Joint Conference on Biometrics (IJCB)*, Abu Dhabi, United Arab Emirates (UAE): IEEE, 2022.

[59] A. Pannatier, R. Picatoste, and F. Fleuret, “Efficient wind speed nowcasting with gpu-accelerated nearest neighbors algorithm,” in *Proceedings of SIAM Data Mining*, Virginia US and Virtual, Apr. 2022.

[60] A. Parvaneh, E. Abbasnejad, D. Teney, R. Haffari, A. van den Hengel, and J. Q. Shi, “Active learning by feature mixing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[61] G. Peronato and J. Kämpf, “Dhgen: Automated generation of district heating network layouts for feasibility studies,” in -, 2022.

[62] F. Piras, E. De Moura Presa, P. Wellig, and M. Liebling, “Local estimation of parametric point spread functions in thermal images via convolutional neural networks,” in *SPIE sensors + imaging, Target and Background Signatures VIII*, vol. 12270, 2022, pp. 1227009 1–8.

[63] A. Prasad, Z.-G. Juan, P. Motlicek, S. S. Sarfjoo, N. Iuliia, O. Ohneiser, and H. Helmke, “Grammar based speaker role identification for air traffic control speech recognition,” in *12th SESAR Innovation Days*, 2022.

[64] A. Prasad, Z.-G. Juan, P. Motlicek, S. S. Sarfjoo, N. Iuliia, and K. Vesely, “Speech and natural language processing technologies for pseudo-pilot simulator,” in *12th SESAR Innovation Days*, 2022.

[65] T. Purohit, I. Ben Mahmoud, B. Vlasenko, and M. Magimai.-Doss, "Comparing supervised and self-supervised embedding for exvo multi-task learning track," in *Proceedings of the ICML Expressive Vocalizations Workshop held in conjunction with the 39th International Conference on Machine Learning*, 2022.

[66] G. Raposo, A. Trajman, and A. Anjos, "Pulmonary tuberculosis screening from radiological signs on chest x-ray images using deep models," in *Union World Conference on Lung Health*, The Union, Nov. 2022.

[67] J. Rozanova, D. Mendes, M. Thayaparan, M. Valentino, and A. Freitas, "Decomposing natural logic inferences for neural nli," in *BlackBoxNLP: Workshop on analyzing and interpreting neural networks for NLP*, 2022.

[68] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, "Are gan-based morphs threatening face recognition?" In *International Conference on Acoustics, Speech and Signal Processing*, May 2022.

[69] E. Sarkar, R. Prasad, and M. Magimai.-Doss, "Unsupervised voice activity detection by modeling source and system information using zero frequency filtering," in *Proceedings of Interspeech*, 2022.

[70] V. Savaliya, A. Bhatnagar, N. Bhavsar, M. Singh, and P. Motlicek, "Innovators@smm4h'22: An ensembles approach for stance and premise classification of covid-19 health mandates tweets," in *International Conference on Computational Linguistics (COLING 2022)*, 2022.

[71] V. Shevchenko, E. Abbasnejad, A. Dick, A. van den Hengel, and D. Teney, "Ebms vs. cl: Exploring self-supervised visual pretraining for visual question answering," in *ArXiv*, 2022.

[72] V. Shevchenko, D. Teney, A. Dick, and A. van den Hengel, "Reasoning over vision and language: Exploring the benefits of supplemental knowledge," in *ArXiv*, 2022.

[73] K. Shinde, T. Ghosal, M. Singh, and O. Bojar, "Automatic minuting: A pipeline method for generating minutes," in *Pacific Asia Conference on Language, Information and Computation (PACLIC 36), In proceedings of ACL Anthology*, 2022.

[74] M. Singh and P. Motlicek, "Idiap submission@lt-edi-acl2022 : Hope speech detection for equality, diversity and inclusion," in *ACL*, 2022.

[75] M. Singh and P. Motlicek, "Idiap submission@lt-edi-acl2022: Detecting signs of depression from social media text," in *ACL*, 2022.

[76] M. Singh and P. Motlicek, "Idiap submission@lt-edi-acl2022: Homophobia/transphobia detection in social media comments," in *ACL Proceedings*, 2022.

[77] Y. Solewicz, N. Cohen, J. Rohdin, S. Madikeri, and H. Cernocky, "Speaker recognition on mono-channel telephony recordings," in *The Speaker and Language Recognition Workshop*, 2022.

[78] S. Srinivas, K. Matoba, H. Lakkaraju, and F. Fleuret, "Efficient training of low-curvature neural networks," in *NeurIPS 2022*, 2022.

[79] N. Tarigopula, S. Tornay, S. Muralidhar, and M. Magimai.-Doss, "Towards accessible sign language learning and assessment," in *ACM International Conference on Multimodal Interaction*, Nov. 2022.

[80] D. Teney, E. Abbasnejad, S. Lucey, and A. van den Hengel, "Evading the simplicity bias: Training a diverse set of models discovers solutions with superior ood generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[81] D. Teney, Y. Lin, S. J. Oh, and E. Abbasnejad, "Id and ood performance are sometimes inversely correlated on real-world datasets," in *ArXiv*, 2022.

[82] B. Ti, Y. Gao, J. Zhao, and S. Calinon, "Imitation of manipulation skills using multiple geometries," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, 2022.

[83] A. Unnervik and S. Marcel, "An anomaly detection approach for backdoored neural networks: Face recognition as a case study," in *21st International Conference of the Biometrics Special Interest Group (BIOSIG 2022)*, Darmstadt, Germany, Sep. 2022.

[84] A. Upadhyay, A. Bhatnagar, N. Bhavsar, M. Singh, and P. Motlicek, “An empirical comparison of semantic similarity methods for analyzing down-streaming automatic minuting task,” in *Pacific Asia Conference on Language, Information and Computation (PACLIC 36), In ACL Anthology Proceedings*, 2022.

[85] A. Upadhyay, N. Bhavsar, A. Bhatnagar, M. Singh, and P. Motlicek, “Automatic summarization for creative writing: Denoising auto-encoder based pipeline method for generating summary of movie scripts,” in *Automatic Summarization for Creative Writing, International Conference on Computational Linguistics (COLING 2022)*, 2022.

[86] M. Valentino, D. Mendes, M. Thayaparan, A. Freitas, and D. Ustalov, “Textgraphs 2022 shared task on natural language premise selection,” in *Proceedings of TextGraphs-16: Graph-based Methods for Natural Language Processing*, 2022.

[87] M. Valentino, M. Thayaparan, and A. Freitas, “Case-based abductive natural language inference,” in *Proceedings of the 29th International Conference on Computational Linguistics*, 2022.

[88] M. Valentino, M. Thayaparan, D. Mendes, and A. Freitas, “Hybrid autoregressive inference for scalable multi-hop explanation regeneration,” in *Thirty-Sixth AAAI Conference on Artificial Intelligence*, 2022.

[89] D. Vallejo-Aldana, A. P. López-Monroy, and E. VILLATORO-TELLO, “Leveraging events sub-categories for violent-events detection in social media,” in *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*, vol. 3202, Sep. 2022.

[90] E. VILLATORO-TELLO, S. Madikeri, P. Motlicek, A. Ganapathiraju, and A. V. Ivanov, “Expanded lattice embeddings for spoken document retrieval on informal meetings,” in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Jul. 2022.

[91] A. Vyas, W.-N. Hsu, M. Auli, and A. Baevski, “On-demand compute reduction with stochastic wav2vec 2.0,” in *Proceedings of Interspeech*, 2022.

[92] J. Wang, T. S. Lembono, S. Kim, S. Calinon, S. Vijayakumar, and S. Tonneau, “Learning to guide online multi-contact receding horizon planning,” in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, 2022.

[93] S. Wazarkar, muskan garg, M. Singh, and O. Bojar, “Multimodality for nlp-centered applications: Resources, advances and frontiers,” in *International Conference on Language Resources and Evaluation (LREC 2022)*, 2022.

[94] S. Yadav, T. Purohit, Z. Mostaani, B. Vlasenko, and M. Magimai.-Doss, “Comparing biosignal and acoustic feature representation for continuous emotion recognition,” in *International Multimodal Sentiment Analysis Workshop and Challenge*, 2022.

[95] M. Zohair, A. Bhatnagar, N. Bhavsar, M. Singh, and P. Motlicek, “Innovators@smm4h’22: An ensembles approach for self-reporting of covid-19 vaccination status tweets,” in *International Conference on Computational Linguistics (COLING 2022)*, 2022.

7 List of Patents

While evaluating the success of technology transfer is a difficult task, invention disclosures and patents are usually key metrics in the field. As a necessary first step to any patent filing, an invention disclosure aims to identify a technology with a certain level of maturity and some promising economic potential.

A patent committee examines each invention disclosure and decides to move forward with a patent filing or not. The committee is composed of the head of technology transfer, two representatives of the direction and one representative of the researchers. Since 2012, 14 patent applications have been filed: 7 patents have been granted, 4 applications are published and pending (1 of them was sold to a company) and 3 application were abandoned. In 2022, the US application of IDIAP-11 has been granted.

- Granted patents

  IDIAP-1 [US 9,689,959 B2] A. Asaei, H. Bourlard, V. Cevher, “Method, apparatus and computer program product for determining the location of a plurality of speech sources”

  IDIAP-2 [US 9,058,541 B2] C. Dubout, F. Fleuret, “Object detection method, object detector and object detection computer program”

  IDIAP-5 [EP3154407] K. A. Funes Mora, J-M. Odobez, “A gaze estimation method and apparatus”

  IDIAP-8 [US 9,973,503 B2] S. Marcel, A. Anjos, P. Abbet, “Method and internet-connected server for reviewing a computer-executable experiment”

  IDIAP-10 [EP 3 691 258 B1] M. Liebling, C. Jacques, “System and method for acquiring images”

  IDIAP-11 [US 11,514,918] | [EP 3 719 679 B1] S. Marcel, V. Krivokuca, “A method for protecting biometric templates, and a system and method for verifying a speaker’s identity”

- Pending applications

  IDIAP-9 [WO 2017/221049 A1] A. Anjos, S. Marcel, “A data-network connected server, a device, a platform and a method for conducting computer-executable experiments”

  globID-1 [WO 2019/150254 A1] L. Sonna Momo, L. Cerqueira Torres, S. Marcel, A. Anjos, M. Liebling, A. Shajkofci, S. Amoos, A. Woeffray, A. Sierro, “Method and device for biometric vascular recognition and/or identification”

  IDIAP-USC-1 [US 2022 0398820 A1] W. Abd-Almageed, L. Spinoulas, M. E. Hussein, D. Geissbühler, S. Marcel, “Multispectral biometrics system”

- Sold applications

  IDIAP-4 [WO 2016/023582 A1] S. Marcel, “A method of detecting a falsified presentation to a vascular recognition system”. This patent has been sold.

- Abandoned applications

  IDIAP-6 [US 2017/0069306 A1] A. Asaei, M. Cernak, H. Bourlard, “Signal processing method and apparatus based on structured sparsity of phonological features”

  IDIAP-9 [US 2019/0180040 A1] A. Anjos, S. Marcel, “A data-network connected server, a device, a platform and a method for conducting computer-executable experiments”

  SAMSUNG-1 [US 2014/0149104 A1, EP 2736042 A1] N-H. Kim, P. Motlicek, P. N. Garner, D. Imseng, J-W. Lee, J-M. Cho, “Apparatus and method for constructing multilingual acoustic model and computer readable recording medium for storing program for performing the method”