ENIA Chair in Artificial Intelligence and Language Technology

Eneko Agirre[1], Olatz Arbelaitz[2], Olatz Arregi[1], Gorka Azkune[1], Arantza Casillas[1], Inma Hernaez[3], Mikel Iruskieta[1], Elena Lazkano[4], Eva Navas[3], German Rigau[1], Roberto Santana[5], Aitor Soroa[1] and Rabih Zbib[5]

[1]HiTZ Basque Center for Language Technology - Ixa NLP Group, University of the Basque Country UPV/EHU
[2]Aldapa, University of the Basque Country UPV/EHU
[3]HiTZ Basque Center for Language Technology - Aholab Signal Processing Laboratory, University of the Basque Country UPV/EHU
[4]RSAIT, University of the Basque Country UPV/EHU
[5]Intelligence Systems Group, University of the Basque Country UPV/EHU
[5]Avature

Abstract

The Chair of Artificial Intelligence and Language Technology has an ambitious program to strengthen leadership in this technology and place the country at the technological forefront. To achieve this, it is supported by two pillars: On the one hand, scientific excellence at the HiTZ research center, Basque Center for Language Technology of the UPV/EHU, in collaboration with the Faculty of Informatics of the UPV/EHU. On the other hand, the company Avature, whose transparent and intentional approach regarding the development of artificial intelligence makes their platform to be chosen by many international companies and government entities to safely deploy their technology in talent-related tasks. The main objective is to reinforce the research leadership with measures to increase the training of experts, the synergy with the industry, the social awareness of the real risks and research into more efficient and unbiased language models. There are four concerted lines of action planned for the next four years: 1) Increase the number of experts through undergraduate, postgraduate and continuous training offer. 2) Technological transfer through a novel collaboration mechanism. 3) Foster research into generative Artificial Intelligence, improving aspects such as energy efficiency and biases. 4) Improve the projection of research to society through a dissemination plan with special emphasis on generative AI, since a society better informed about the capabilities of AI will be able to adopt and enhance the benefits, and in turn identify and mitigate the new challenges it entails.

Keywords

Artificial Intelligence, Language Technology, Natural Language Processing, Outreach, Talent, Innovation

1. Introduction

The ability to communicate through natural language is a key aspect of intelligence. Language Technology is the field within Artificial Intelligence that focuses on the development of algorithms and systems that can understand, interpret and even generate human language. This field is leading a revolution within AI, with applications that are changing society, such as automatic translation, voice conversational systems or generative systems like GPT or Gemini.

The Chair of Artificial Intelligence and Language Technology has an ambitious program to strengthen leadership in this technology and place the country at the technological forefront. To achieve this, it is supported by two pillars: On the one hand, scientific excellence at the HiTZ research center, Basque Center for Language Technology of the UPV/EHU, in collaboration with the

Faculty of Informatics of the UPV/EHU. On the other hand, the company Avature, whose transparent and intentional approach regarding the development of artificial intelligence makes their platform to be chosen by many international companies and government entities to safely deploy their technology in talent-related tasks.

The Chair is part of the National Artificial Intelligence Strategy (ENIA), part of the Digital Spain 2026 agenda, and has been selected in the call of the Ministry of Digital Transformation to finance university chairs dedicated to research, dissemination, teaching and innovation on Artificial Intelligence. The chair includes a procedure to incorporate additional interested companies.

The **objective** is to reinforce the research leadership with measures to increase the training of experts, the synergy with the industry, the social awareness of the real risks and research into more efficient and unbiased language models.

The chair was created on February 22, 2024 and has a duration of four years. The chair has identified four concerted lines of action:

- Increase the number of experts through undergraduate and postgraduate studies, as well as the

continuous training of technicians, managers and social agents in this technology.

- Transfer between the chair and companies through a novel collaboration mechanism that seeks the synergy between them, detecting the most promising areas.
- Foster research into generative Artificial Intelligence, improving aspects such as energy efficiency and biases.
- Improve the projection of research to society through a dissemination plan with special emphasis on generative AI, since a society better informed about the capabilities of AI will be able to adopt and enhance the benefits, and in turn identify and mitigate the new challenges it entails.

Before looking at each of the lines of action, we present the motivation for this chair, and the needs detected.

2. Motivation

The ability to communicate through natural language is a key aspect of human communication and is essential to effectively transmit information. Natural language processing (NLP) and language technology (TL) are fields within Artificial Intelligence (AI) that focus on the development of algorithms and systems that can understand and interpret human language. NLP is an important area of AI with numerous applications, particularly in the context of today's digital transformation. Language is and must be at the center of our efforts to develop AI, and vice versa. In fact, currently TL is arguably the most innovative field of AI with rapidly growing economic impact. Both AI and NLP are undoubtedly expanding fields that generate an immense volume of business and qualified jobs. The global NLP market is estimated to be worth $13.5 billion in 2021 and is believed to reach an expected value of $91 billion by 2030, growing at a CAGR of 27% during the forecast period (2022-2030). These figures indicate that the return on investment will be enormous. For example, OpenAI, owner of ChatGPT, estimates one billion dollars in revenue by 2024. It is therefore necessary to guarantee competitiveness in this area, through technological surveillance, training, dissemination, transfer and cutting-edge research, guaranteeing strengthening the entire TL ecosystem of the country.

In recent years the emergence of new and powerful Deep Learning (DL) techniques are revolutionizing NLP. Neural machine translation systems or language models such as ChatGPT or Bard allow developing applications that were unthinkable just a few years ago. As a result, leading technology companies such as Google, Facebook, Microsoft or Amazon have integrated large neural language models pre-trained with huge amounts of data into their systems. Each of these models can be used on many tasks, such as search engines, chatbots, summary creation, translation or computer program generation.

Motivation of the Chair

The NLP community is contributing to the emergence of disruptive new DL techniques and tools that are revolutionizing the approach to language technology LT tasks. The typical way to implement NLP solutions has moved from a pipeline based methodology, to architectures based on complex neural networks trained with vast amounts of data. This rapid progress in NLP has been possible because of the confluence of four different research trends: 1) mature deep learning technology, 2) large amounts of data (and for TL, large and diverse multilingual textual data), 3) increase in high-performance computing power in the form of Graphic Processing Units (GPUs), and 4) application of simple but effective self-learning and transfer learning approaches using Transformers [1, 2, 3, 4].

Thanks to these recent advancements, the NLP community is currently engaged in a paradigm shift with the production and exploitation of large, pre-trained transformer-based language models [5, 6]. As a result, various IT corporations have started deploying large pre-trained neural language models in production. For instance, Google and Microsoft have integrated them in their search engines. Compared to the previous state of the art, the results are so good that systems are claimed to obtain human-level performance in laboratory benchmarks when testing some difficult language understanding tasks.

Due to the great results obtained, nowadays there is a tendency to build increasingly larger models. For instance, GPT-3 contains 175 billions of parameters and was trained on 570 gigabytes of text, with a cost estimated between ten and twenty million USD. In comparison, its predecessor GPT-2 was 100 times smaller, with 1.5 billions of parameters [7]. This upscaling leads to surprising behavior: GPT-3 is able to resolve tasks for which they have not been previously trained, just by providing them with very few training examples. This behavior was not observed in the much smaller GPT-2 model. There are works that indicate that this behavior depends directly on the scale [8], so that, from a specific size, models show emerging abilities that allow them to undertake tasks for which they have not been taught.

Large language models obtain remarkable results but are extremely costly to train and develop, both financially, due to the cost of hardware and electricity or cloud computing time, and environmentally, due to the carbon footprint required to fuel modern servers with multiple Graphics Processing Unit (GPU) hardware. This also means that only a limited number of organisations with

abundant resources in terms of funding, computing capabilities, NLP experts and data can currently afford to develop and deploy such models, thus aggravating the lack of technology sovereignty in most linguistic communities.

There are also worrying shortcomings in the text corpora used to train these anglo-centric models, such as the predominance of harmful stereotypes or the lack of representation of less-resource languages (which in fact represent the vast majority of languages, and include Catalan, Basque and Galician). Language models are known to be biased in several respects, and they also lack explainability, that is, the decisions made by these models cannot be scrutinized, due to their black box nature.

On the other hand, we must be aware of the risks of the massive use of AI&TL in society. For example, this technology usually works much better in certain languages with large number of speakers, while the vast majority of the languages run the risk of being excluded from their benefits, thus losing attractiveness and generating a worrying sociolinguistic impact. On the other hand, biases may have harmful effects on society. In another example, neural translation and text generation systems may produce worrying changes in the information flow, such as the spread of fake news or plagiarism. These risks are also worry for companies that use TL, as their products are sensitive to these risks, which can lead to suboptimal results and, in some cases, serious reputation problems.

Needs

The necessity for further exploration and research in the field of AI&TL is evident. Modern AI&TL is mostly based on deep neural networks, a key-enabling technology that has shown enormous potential in the productive and social fabric. In this context, it is paramount for the university and society in general to react, and to avoid being mere passive spectators of these disruptive changes.

The University of the Basque Country already works on AI&TL within its educational, transfer, and research activities. However, the Chair aims to go a further step and become an institutional unit researching on AI&TL that systematically helps generating an explicit reference point inside the university itself, therefore contributing to visualize, promote, energize and join all efforts towards this direction, both in the university and in society.

Besides, the Chair should also undertake dissemination and dynamization tasks in society. It has to serve as a connection agent that facilitates a knowledge flow from academy towards society, and in turn feed by its needs and concerns.

From a social point of view, there are urgent needs regarding education. Bachelor's and master's students are often unaware of the specific applications of the techniques they study, and they could greatly benefit by collaborating with Chair's entities when developing their final degree or master's thesis. Likewise, the lack of computational resources available to students limit the scope of their work. On the other hand, technical staff in companies, research/technological centers and administrations are demanding continuous education programmes to widen their knowledge base and enhance their technical skills in these latest advances. This is applicable not only to software engineers, but also to technicians from other areas who now see how AI&TL are causing disruption in their domains. Likewise, the managing staff in companies, technology/research centers, administrations are often unaware of the opportunities and threats that AI&TL technologies pose.

It is also necessary to devise a good dissemination strategy to raise the conscience about AI&TL in society, both their possible benefits but also its harms. New technologies tend to generate opinions that are not always well-founded, both on the exaggeratedly optimistic side (AI&TL like ChatGPT is an intelligent being like humans), and on the catastrophic side (AI&TL is so powerful that it will leave screenwriters out of work). These exaggerations can alter the perception that general public has regarding subtle effects, such as biases in automatic translation or the real impact of text generation systems in classrooms, which are therefore not known to their full extent.

Regarding transfer as well as industry and administrations needs, there is a general unawareness of the possibilities of AI&TL, and agents and practitioners need a framework where they can share their needs and capabilities, and hence be able to design processes and applications that do not exist today.

Finally, society requires research that takes the needs of companies and administrations into consideration, incorporating realistic scenarios and data, or building systems that can be deployed in environments where the computational requirements are constrained. The Chair will constantly monitor and detect new needs, and adapt its actions accordingly.

Social Relevance

A society that is properly informed about the capabilities of AI&TL will be able to adopt and enhance its benefits while identifying and mitigating the new challenges they may pose. Administration directors who are properly trained and informed about AI&TL will be able to make effective decisions that improve society in general. In turn, a forum where companies developing AI&TL, consumer companies and university researchers share their potentials and needs will allow for more effective transfer, research oriented to society, and ultimately a synergy and mutual benefit.

3. Training

The training area of the Chair contemplates two lines of work, training of undergraduate and master students, and continuous training.

Regarding the training of undergraduate and master students, the Chair will serve to reinforce the training of students through their final degree projects. To this end, Final Degree Projects and Master's Degree Projects will be defined in the subject of the Chair in bachelor's and master's degrees related to AI&LT. The participation of the entities of the Chair will be encouraged in the definition of these projects and in their co-direction. Likewise, internships will be defined and carried out in the companies and institutions associated to the Chair, as well as in other companies and institutions with relevant activity in LT. In this way, students will have the opportunity to to broaden and deepen their knowledge and skills with more practical aspects, which sometimes left out the usual regulated teaching. In addition, students will have the opportunity to attend the continuing training courses that will be offered to personnel outside the university, as well as the dissemination, research and transfer activities detailed in the following sections.

Regarding continuous training, it is planned to offer, together with the entities of the Chair, an own title of expert in AI&LT that aims to cover the training needs in the latest advances in the area. This degree will include basic training courses, together with other more technical courses on the most current topics in LT such as deep learning or large language models, as well as a third block of complementary courses on more diverse and specific topics, such as AI regulation, opportunities in business management and/or the explainability of models, among others. These courses will be offered also as independent modules, which can be chosen according to the needs of each one, and are aimed either at technically trained people who need to deepen and update their knowledge, or at company managers and administration managers who want to know the opportunities that LT can offer them in their work environment. The curriculum will be designed according to the real needs of the industry and institutions. In order to reach as many people as possible, a fully online version will be offered, as well as a blended learning version.

4. Research Transfer

The chair will apply the following actions for the exploitation of research, training and dissemination results with measures aimed at increasing the impact of the results on the productive/scientific, social and economic fabric. (1) Exploitation of the two lines of research, and search for a US patent for each of them. (2) Organization of an annual event to promote the exploitation of the results of the chair and AI&LT technology. (3) The chair will be established as a framework for the chair's entities to share their needs and capabilities in order to design disruptive processes and applications, beyond what is possible to do today. The latest advances obtained in AI&LT by the participants of the chair will be presented. Several collaborative projects will be chosen by the participants of the chair, offering the servers owned by the chair when possible.

5. Research

The following activities will be encouraged: (1) Internally, a conversation will be started to study the possible orientation of existing research to the areas of interest of the chair's participants. This conversation has already taken place with Avature, with which two lines of work have been selected: a) Biases and explainability: user empowerment. b) Green LT: reduce the computational cost in job recommendation algorithms. (2) Two thesis projects will be defined and selected from among the research lines of interest to the chair. These theses will have an applied aspect, in the sense of focusing on real problems that concern the business entities part of the chair. (3) Externally, the chair will participate in research groups and collaborate with other groups with common objectives. On the other hand, open scientific events will be organized on the subject area of the chair (see Outreach).

6. Outreach

The Chair will organize the following tasks: (1) Dissemination events to society, including participation in science weeks and similar events. Participation in the media such as radio, social media, etc. Top-level academic events, including the organization of a leading IA&TL conference, SCIE Class 1. (2) Dissemination events with the participation of companies and public administrations with the aim of disseminating know-how transfer. In addition, open events will be organized that coordinate the collaboration of different types of participants: new techniques researchers, new products developers, corporate software integrators, AI&TL consumers, end users and administration. (3) Scientific dissemination events, such as monthly webinars on innovative research and transfer topics, as well as social or educational topics about AI and TL that arouse interest in the society.

7. Research Groups participating in the project

The chair is closely linked to the HiTZ Basque Center for Language Technology[1] and the Computer Science Faculty, both in the University of the Basque Country UPV/EHU. It comprises the research groups below. The other pillar of the chair is Avature, also described below.

Aholab Signal Processing Laboratory. Aholab[2] is the short name of the Signal Processing Laboratory of the University of the Basque Country (UPV/EHU). The laboratory is located in Bilbao. We are a university research team and focus our research in the areas of Text to Speech Conversion, Speech and Speaker Recognition, and Speech Processing in general. Since 2005 we are a recognized research group of the Basque Research Network. The laboratory is part of the Basque Center for Language Technology (HiTZ) and the Department of Communications Engineering of the Faculty of Engineering of Bilbao (EIB).

Ixa NLP Group. Ixa[3] is a research group from the University of the Basque Country (UPV/EHU) that works in all areas of Natural Language Processing. Ixa is a multidisciplinary group with more than 25 years of experience, comprising computer scientists, linguists and other disciplines. The group is based on the Computer Science Faculty in San Sebastian and the Languages and Computer Systems department, but many members belong to other faculties and departments of the UPV/EHU. The group is part of the Basque Center for Language Technology (HiTZ).

Aldapa group. ALDAPA [4](ALgorithms, DAta mining and PArallelism) is a research group from the University of the Basque Country (UPV/EHU) mainly based on the Computer Architecture and Technology department of the Computer Science Faculty but it also includes staff from other engineering schools in the UPV/EHU. The group has more than 25 years of experience using, proposing and adapting machine learning algorithms to solve real world problems, having currently two main focuses: artificial intelligence and physiological computing in health, specifically in diagnosis and prevention of nervous system diseases (Parkinson's, Alzheimer's, stress, epilepsy), and, fair and explainable machine learning models.

ISG Intelligent Systems Group. The Intelligent Systems Group (ISG)[5] focuses its research in the areas of machine learning, combinatorial optimization and high performance computing, where it has an extensive research experience. ISG is one of the UPV/EHU groups with a consolidated category "A" group in the latest call for research groups of the Basque Government. Its vocation for high-quality basic research is reflected in the more than 200 publications in international journals with impact factor, and in the 16 doctoral theses defended during the last 5 years.

RSAIT Robotika eta Sistema Adimendunen Taldea. The Robotics and Autonomous Systems research group develops its activity at the Computer Science Faculty of the UPV/EHU since 2005. Nowadays, its research efforts focus in two main areas: the development and application of machine learning techniques, mainly for activity recognition in videos; and the semantic based co-speech gesture generation in social robots. It is a consolidated A group according to the Basque Government and has a long participatory tradition in cultural events, disseminating the state of development of both, AI and robotics.

Avature. Avature[6] is a highly configurable software-as-a-service (SaaS) platform for enterprise Human Capital Management (HCM), focusing on talent acquisition and management. It is the leading provider of CRM and ATS technology for human resources worldwide. Founded by Dimitri Boylan, co-founder and former CEO of HotJobs.com, Avature brings consumer-driven internet technology and digital innovation to human resources departments. Avature solutions are used by 110 Fortune 500 companies in more than 164 countries and in 32 languages. It delivers its services from its private cloud located in data centers in the US, Europe and Asia. Avature has offices in Buenos Aires, London, Madrid, Melbourne, Munich, New York, Paris, Shenzhen and Virginia.

Acknowledgments

---

[1] https://hitz.ehu.eus
[2] https://aholab.ehu.eus/
[3] htpps://ixa.ehu.eus
[4] https://www.aldapa.eus

[5] http://www.sc.ehu.es/ccwbayes/
[6] www.avature.net

References

2206.07682.

[1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[2] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, L. Zettlemoyer, Multilingual denoising pre-training for neural machine translation, Transactions of the Association for Computational Linguistics 8 (2020) 726–742. URL: https://aclanthology.org/2020.tacl-1.47. doi:10.1162/tacl_a_00343.

[3] J. Pfeiffer, I. Vulić, I. Gurevych, S. Ruder, MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 7654–7673.

[4] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Q. Liu, D. Schlangen (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https://aclanthology.org/2020.emnlp-demos.6. doi:10.18653/v1/2020.emnlp-demos.6.

[5] H. Li, Language models: past, present, and future, Commun. ACM 65 (2022) 56–63. URL: https://doi.org/10.1145/3490443. doi:10.1145/3490443.

[6] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, D. Roth, Recent advances in natural language processing via large pre-trained language models: A survey, ACM Comput. Surv. 56 (2023). URL: https://doi.org/10.1145/3605943. doi:10.1145/3605943.

[7] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (2019) 9.

[8] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, W. Fedus, Emergent abilities of large language models, 2022. URL: https://arxiv.org/abs/2206.07682. doi:10.48550/ARXIV.