International Chair Generative AI: Challenges & Risks

USP CEU - Deloitte

With the collaboration of:

GENERATIVE AI: OPPORTUNITIES & RISKS SUMMIT

Description and objective

Generative Artificial Intelligence (IA) stands as a technology whose rapid growth presents both significant opportunities, but also challenges for our societies. On the one hand, its capacity to generate realistic, new content has the potential to enhance creativity and spark innovation across many industries, while also helping advance scientific research. On the other hand, its deployment within systems operating in real-world settings conveys risks and questions that need to be addressed now, including but not limited to the possible impact on the job market, the spread of misinformation, the creation of harmful content, or the systematic reproduction of data-inherited biases.

The aim of this international Summit is to bring together a team of interdisciplinary researchers and industry leaders from some of the most influential institutions in the world to address the challenges and risks that Generative AI poses, and to publicly share their conclusions.

What?

International Hight-level Experts Workshops (closed-door event without audience, invitation only)

Format

Closed to public. The Summit aims to work, behind closed doors, with high-level experts to discuss and draw relevant conclusions on the topics analysed.

Where?

Real Colegio Complutense
Cambridge, Harvard Main Campus area (TBC)
26 Trowbridge St, Cambridge, MA 02138, USA

Invited experts:

Idoia Salazar

International Director 'Generative AI: Risk and Challenges' Chair
President OdiseIA
USP-CEU
Physically confirmed.
https://www.linkedin.com/in/idoiasalazar/

Richard Benjamins

CEO
Former Chief Responsible AI Officer
OdiseIA
Physically confirmed.
https://www.linkedin.com/in/richard-benjamins/

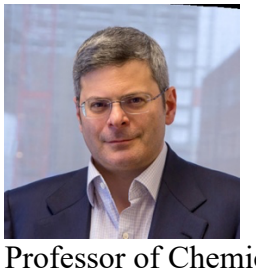
Ricardo Kleinlein

Postdoctoral Research Fellow
Harvard Medical School
https://www.linkedin.com/in/ricardo-kleinlein/
Physically confirmed.

Miguel Ángel Armengol

LCP
Massachusetts Institute of Technology
https://www.linkedin.com/in/miguelangelarmengol/
Online participation.

Bernhardt L Trout



Professor of Chemical Engineering

MIT and Member of the Advisory Board at Hedgeness
Physically confirmed.

Matilda Dorotic

Visiting scholar at Harvard University
Visiting scholar at Massachusetts Institute of Technology
Associate Professor at Department of marketing, BI Norwegian Business School
To be confirmed.

Leo Anthony Celi

Associate Professor Of Medicine, Part-Time at Harvard Medical School
https://www.linkedin.com/in/leo-anthony-celi-b25131/

Elizabeth Renieris

Author of "Beyond Data" (MIT Press), Privacy & AI Expert
https://www.linkedin.com/in/elizabethrenieris/
Physically confirmed.

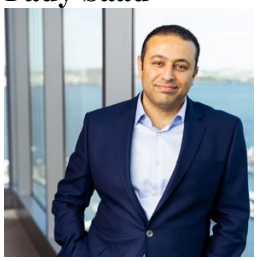
Allison Ryder

Program Director, Research
MIT Sloan Management Review
https://www.linkedin.com/in/allisonryder/
Physically confirmed.

Stefaan Verhulst

Co-Founder, Chief Research and Development Officer and Director Data Program
The GovLab, New York University
https://www.linkedin.com/in/stefaan-verhulst/
Physically confirmed.

Fady Saad



General Partner
Cybernetix Ventures
MassRobotics Co-Founder
https://www.linkedin.com/in/fady-saad-0a1972/
Physically confirmed.

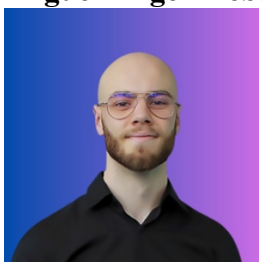
Esteban Moro

Professor of Physics & Network Science Institute
Northeastern University
https://www.linkedin.com/in/esteban-moro/
Physically confirmed.

Julio Gonzalez Pacheco

Director
NTT Data
https://www.linkedin.com/in/jaglezpac/
Physically confirmed.

Miguel Ángel Liébanas



CEO
Human Trends
https://www.linkedin.com/in/miguel-%C3%A1ngel-li%C3%A9banas-559182185/
Online Participation.

Outputs:

1. The primary objective of this international Summit is the elaboration of a specific and grounded report documenting the discussions and conclusions of the event regarding the risks, challenges and opportunities posed by Generative AI-based tools in the healthcare sector.
2. Our secondary objective is to identify those aspects of current AI regulations that, when referring to Generative AI in the health field, may be open for refinement and improvement, both in the short-, mid- and long-term.
3. A research action plan shall be proposed, so further work can be continued through the professional network established during the event.
4. The conclusions of this Summit will be published and shared through collaborating entities (selected media), social media and the Chair's website to disseminate results to the general public.

Summit conducted using Delphi Methodology

1. Pre-Summit Survey: We´ll be distributing an initial survey to gather baseline opinions on the main topics (risks, challenges, opportunities) from all participants before the summit.

2. Form Expert Panels: Organise attendees into expert panels by specialty area. Each panel focuses on specific topics, working independently and iteratively to refine their findings.

3. Implement Iterative Feedback: After initial panel discussions, circulate the summarised findings among all participants. Conduct multiple feedback rounds, allowing for the revision of opinions based on group inputs.

4. Hold a Final Consensus Meeting: At the summit's conclusion, discuss the latest feedback in a plenary session to make any necessary adjustments and achieve final consensus on the report.

5. Validate Post-Summit: Send a draft of the consensus report to all participants for a final review and validation before publication, ensuring the report accurately reflects the collective agreement and insights.

TOPICS

Ethical risks and challenges of Generativa AI

Generative Artificial Intelligence (AI) technologies hold immense promise for revolutionizing various sectors. However, alongside the potential benefits come ethical risks and challenges that demand careful consideration.

In the particular case of healthcare applications involving Generative AI, one of the most important challenges nowadays is to leverage these tools to facilitate equitable and accessible healthcare to our citizens. To reach that goal, a crucial matter is to understand how discrimination is generated in healthcare and medicine. However, it is essential to recognize that discrimination encompasses disparities rooted in the social determinants of health. These determinants, including factors such as socioeconomic status, education level, and access to healthcare, profoundly influence health outcomes and healthcare experiences, and ultimately will become labels in datasets. Those within our societies who are underrepresented or marginalized may have their voices and experiences overlooked. This underrepresentation poses a significant risk of perpetuating existing disparities in healthcare.

Additionally, another significant pitfall in utilizing Generative AI tools for processing highly sensitive and private data such as medical records, lies in the realm of data privacy and security. Often, the best-performing models are owned by large corporations and accessible solely through APIs. Entrusting such sensitive information to external entities raises profound concerns regarding data governance, confidentiality, and potential breaches.

These are but two instances of ethical risks and challenges that the use of Generative AI encompasses today.

Cultural differences, perception, and the impact of Generative AI

Human cognition is inherently shaped by biases—psychological shortcuts that aid in quick decision-making and efficient processing of information. While these biases serve a crucial function in navigating the complexities of the world, they can also harbor negative consequences. Generative Artificial Intelligence (AI) possesses the capability to exploit these biases, not solely for benevolent purposes. The potential for Generative AI

to leverage biases raises significant ethical concerns, particularly in contexts such as healthcare.

For instance, European regulations concerning AI mandate the inclusion of interpretability tools within algorithms. However, this requirement poses considerable technical challenges, as there is no universally applicable interpretability tool. This aspect of the regulation is particularly weak when applied to Generative AI. The process of generating explanations for AI model responses can be manipulated to exploit known biases, thus exacerbating them with potentially severe consequences. For instance, the deployment of systems that reinforce the tendency for overreliance in clinical practice may worsen health outcomes rather than improving them (Bussone, Stumpf, and O'Sullivan, 2015; Buçinca, Malaya, and Gajos, 2021).

Addressing these challenges requires a multifaceted approach that prioritizes ethical considerations and technological innovation. Proactive measures must be taken to mitigate the risks associated with bias exploitation in Generative AI systems.

AI For Good

While the application of Generative AI for healthcare is a type of AI for Good, there are many other potential applications that are less well developed. Think about Generative AI for solving the SDGs such as hunger, poverty, inequality, climate change, clean water, etc. The large majority of investments in Generative AI today are motivated by profit. The result of this is that areas of application with no direct and clear economic return, such as several SDGs, are experimenting little activity, and are mostly funded in a philanthropic way.
This line of work will analyse what GenAI for Good applications there could be and will come up with policy recommendations to foster their development beyond philanthropy.

Use of Generative AI in health

In the realm of healthcare, the application of Generative AI presents unique challenges and opportunities that distinguish it from other domains. Unlike sectors where data primarily revolves around consumer behavior or preferences, healthcare data encompasses highly sensitive and personal information, such as medical records and patient histories.

The implications of utilizing Generative AI in medicine extend beyond efficiency and convenience to directly impact patient care, treatment outcomes, and overall well-being. Moreover, the criticality of accuracy and reliability in healthcare decision-making necessitates a higher standard of scrutiny for Generative AI systems. Errors or biases in AI-generated recommendations or diagnoses can have profound consequences for patient safety and trust in the healthcare system.

Furthermore, the interdisciplinary nature of healthcare requires collaboration between diverse stakeholders, including clinicians, researchers, ethicists, and policymakers. Integrating Generative AI into medical practice necessitates not only technical expertise but also a nuanced understanding of ethical, legal, and social implications. This

interdisciplinary approach is essential for addressing complex healthcare challenges and fostering responsible innovation in the field.

Additionally, healthcare regulations and standards are rigorously enforced to protect patient rights, privacy, and confidentiality. The introduction of Generative AI into medical practice must adhere to these stringent regulatory frameworks while balancing innovation and patient safety. Compliance with regulatory requirements adds an additional layer of complexity to the development and deployment of Generative AI solutions in healthcare settings.

In the pursuit of integrating Generative AI into healthcare, numerous complex challenges emerge, namely:

- What are the potential ramifications of disregarding the influence of social determinants of health when incorporating Generative AI tools into healthcare systems, potentially perpetuating preexisting disparities and inequalities.
- How can we effectively communicate to stakeholders, developers, and pertinent parties within healthcare the inadequacy of solely relying on statistical accuracy metrics for assessing Generative AI models, and instead, underscore the necessity of understanding the whole data life cycle and its effect on the evaluation of these models.
- What foreseeable impacts might arise from the integration of automatically-generated synthetic data within healthcare settings.
- How might the introduction of Generative AI in healthcare impact patient trust and perceptions of healthcare providers, and what strategies can be employed to mitigate potential concerns and foster transparency and trust?
- What ethical considerations must be taken into account when deploying Generative AI in healthcare, particularly regarding patient consent, privacy rights, and the potential for unintended biases or discrimination.
- How can regulatory bodies adapt to the evolving landscape of Generative AI in healthcare to ensure patient safety, data privacy, and equitable access to innovative technologies while fostering innovation and advancement in medical care.

AGENDA

7th MAY: Introduction to the Summit & Definition of Research Questions

9-9:30 AM	Reception and Networking
9:30 - 10 AM	Words form the Spain Consul in Boston. D.ª Ana Durán de la Colina
10:00 - 10:15 AM	Introduction to the Summit and and brief presentations by the participating experts
10:15 - 10:30 AM	Presentation of the topics to discuss
10:30 - 11:00 AM	Coffee Break

11:00 - 12:30 PM	Panel Discussion 1 - Ethical Risks of Generative AI
12:30 - 1:30 PM	Lunch
1:30 - 3:00 PM	Panel Discussion 2 - Cultural differences, perception and implications
3:00 - 3:30 PM	Coffee Break
3:30 - 5:00 PM	Panel Discussion3 - AI for Good
5:00 - 5:30 PM	Day 1 Closing Remarks

8th MAY: Focused Research Questions

9:00-9:30 AM	Reception & Networking breakfast
9:30 - 9:45 AM	Recap of Day 1
9:45 - 10:15 AM	Introduction to existing issues with GenAI in healthcare
10:15 - 11:15 AM	Breakout Sessions (I)
11:15 - 11:35 AM	Coffee break
11:35 - 12:30 PM	Breakout Sessions (II)
12:30 - 1:30 PM	Lunch
1:30 - 3:00 PM	Report back & Discussion
3:00 - 3:30 PM	Coffee break
3:30 - 4:30 PM	Closing Session - Action Plan & Next Steps
4:30 - 5:00 PM	Summit Closing Remarks

BACKGROUND INFORMATION ON THE INTERNATIONAL CHAIR ON GENERATIVE AI

The International 'Generative AI: Challenges & Risks Chair', is the product of a public, competitive call from the Secretary of State for Digitalization and Artificial Intelligence from the Government of Spain. The main partner are this Secretary of State and the company Deloitte.

Executive summary of the project

Artificial Intelligence is not the future, it is the present. According to the latest McKinsey study, its adoption by organisations around the world has more than doubled since 1970 and is expected to continue to grow at an exponential rate. The technology itself continues to develop at a rapid pace, mainly driven by large companies such as OpenAI and Google. It is difficult to predict its evolution in the medium term and the impact it will have, both professionally and socially. However, one thing is clear: the impact will be drastic. In other words, it will have a significant impact on practically all areas of knowledge and on many facets of our daily lives. The advantages are many, but there are undoubtedly also ethical and legal challenges and risks. We are still only able to glimpse a small part of the great impact that we will experience in the coming years, thanks to the introduction of tools such as ChatGPT or Dall-e. Now is the time to prepare. To anticipate in order to predict possible cases in the near future, as well as to study latent cases. It is now when we must develop tools (why not also AI tools) that will help us with this purpose of

achieving a Reliable AI. And also, it is time to prepare society, in general, for the impact of this technology. A society that still harbours many prejudices derived from science fiction. People, of all ages and conditions, who, in general, lack the necessary knowledge to successfully face the world of data and AI that we are living in. Research, development and dissemination. These are the foundations of the International Chair – "Generative AI: Challenges and Risks" that the CEU San Pablo University intends to create, together with Deloitte and other universities and national and international companies.

Through this body, exhaustive and realistic research will be carried out into the challenges and risks of Generative AI. It will analyse concrete use cases and work on plausible short-medium term predictions to prevent negative or unwanted impacts. The fundamental research consists of two main areas:

- Research to enhance the ethical use of Generative AI. This section will analyse and draw conclusions relating to:
  - Ethical and social impact of Generative AI.
  - Short-medium term predictive analysis of the ethical and regulatory challenges of Generative AI.
  - Importance of cultural, geographical, and temporal perspectives in addressing the ethical and social challenges of Generative AI.
  - Incidence and prevention of discrimination in Generative AI systems.
  - Social study on the incidence of Generative AI on a professional and personal level.
  - Elaboration of a real and grounded strategy to achieve the objectives of the Spanish Government's Digital Rights Charter and the European Digital Rights and Principles, as far as the impact of Generative AI is concerned.
  - Impact of Generative AI in the Metaverse: challenges and risks. Detection and analysis of its prevention.

On the other hand, and always bearing in mind our idea of carrying out research applicable to the changing reality we are living, we also combine our request for fundamental research with another one related to experimental development focused on the two areas developed in our Fundamental Research Centre within the Chair. Experimental technical developments will be carried out, open access, specifically for:

- Prediction, in the short-medium term of use cases related to the social-ethical impact of generative AI.
- Tools to supervise AI systems and 'certify' their correct use on the basis of European regulations.
- Specific tools for detecting different types of biases.
- Tools for risk prevention, at sectoral level, regarding the ethical and social impact of generative AI.
- Trend prediction tools.

In addition, the Chair integrates an **ambitious dissemination plan**, integrated in a third application, with two distinct aspects:

- **Dissemination actions for the general public:**
  - On the one hand, we will work directly with the media (RTVE Institute, EFE Agency, El País...) in the analysis of the use cases presented (e.g. ChatGPT, and other generative AI...), in order to provide the public with reliable information (from experts in the field) on these issues. The link between the management of this Chair and the Observatory of the Social and Ethical Impact of Artificial Intelligence (OdiseIA), currently made up of more than 200 interdisciplinary experts (national and international) in this field, is particularly relevant to this issue.
  - Work will also be carried out to detect trends in the public's perception of generative AI. And, together with the collaborating entities belonging to the field of communication, specific strategies will be developed to mitigate possible risks of misinformation in this regard.
  - A specific action plan for senior citizens will be implemented. Tailored seminars and workshops will be offered to understand the real impact of this technology.
  - A specific action plan will be developed to encourage the participation of women in AI management and development teams.
  - A specific action plan will be developed to encourage an interdisciplinary approach to these technologies: from the professional and educational spheres.
  - A proposal will be developed for primary and secondary schools to include specific knowledge on the impact of AI in formal education. This proposal will be tested in CEU schools.

- **Actions to disseminate scientific knowledge:**
  - Specific international congresses on key research topics.
  - Expert round tables.
  - Seminars and webinars.
  - Meetings, behind closed doors, with international experts to get to know different perspectives.

By approaching dissemination from the different fronts defined above, the aim is to forge an important reliable basis for the correct development and implementation of this technology.