



ISSUE

AI REGRESSION OR PROGRESS?



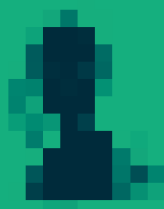
-00:36:12:41:08



-00:22:12:41:34



-00:09:12:41:46



-00:06:33:00:45



-00:00:57:41:22



-00:00:42:41:55



00:00:00:00:00



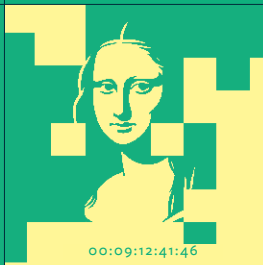
00:00:42:41:55



00:00:57:41:22



00:06:33:00:45



00:09:12:41:46



00:36:12:41:08



TABLE OF CONTENTS

Guest editors
P. 5

Introduction to the report
P. 6

Substitution
P. 8

GANs and Generative AI:
the Challenge of Disinformation
and Cognitive Warfare
P. 14

AI and the New Bipolar
Geopolitical Landscape
P. 19

The Use of AI in Recruitment
and the Risks of Mirages
P. 24

Soundscapes and AI:
Ecological Issues
P. 29

AI, Historical Research and the
Maturing Relationship between
History and Computer Sciences
P. 34

The Ambivalence of Professional
Journalism Faced with AI's
“Surveillance Capitalism”
and Influence
P. 40

The Uses of Large Language Models
in the Social Sciences
P. 44

GUEST EDITORS



Camille Salinesi

is a professor of computer sciences at Paris 1 Panthéon-Sorbonne University, affiliated with the Centre for Research in Computer Science (CRI) and co-director of the Artificial Intelligence Observatory. He has published numerous articles on the design of socio-technical and cyber-physical systems, particularly using AI techniques, and on the problems of designing applied AI systems. He is co-author of the *Karlskrona Manifesto for Sustainable Design* and the future-predicting keynote address “*One day, information systems will design themselves*”.



Célia Zolynski

is an associate professor of private law at the Sorbonne Law School of Paris 1 Panthéon-Sorbonne University, where she co-directs the Sorbonne's Department of Research in Intangible Law (IRJS-DreDis) and the Artificial Intelligence Observatory. She also directs the Master 2 degree in Creative and Digital Law and was a member of the National Steering Committee on Ethics and Digital Technology (CNPEN) from 2019 to 2024. She is currently a qualified expert within France's National Consultative Commission on Human Rights (CNCDH) and the Higher Council for Literary and Artistic Property (CSPLA), as well as being a member of the CNIL's Prospective Committee. Her research and teaching activities focus on digital law, intellectual property law, market law and fundamental freedoms.



Stéphane Lamassé

is a senior lecturer in medieval history at Paris 1 Panthéon-Sorbonne University. He specialises in the history of medieval science in the field of mathematics and has focused part of his research on historical processing of digital information, as well as on methods that can be used by historians in terms of data analysis.



PRESENTATION OF THE TRANSLATOR

NICHOLAS SOWELS

Nicholas Sowels is a Senior Lecturer in English for economics at the Département des langues, where he has taught since the late 1990s. His present research areas include economic and public policy in the United Kingdom, Brexit and finance, as well as poverty and inequality in the UK. He also works as a freelance translator and editor of texts written in English by non-native speakers.

It took only a few months to realise that artificial intelligence (AI) had left the labs and entered our daily lives. Since the emergence of ChatGPT at the end of 2022, anyone can experience it: you can ask a machine to summarise a text, write an email, correct an assignment, generate an image, simulate a voice or compose a piece of music. Previously, GPS navigation assistance, diagnostic assistance, product and content recommendations, fingerprint recognition and weather forecasts were already making extensive use of AI. But now the technical wonder of AI is here more generally, and new uses are rapidly developing.



However, since AI seems to have mastered language, our fascination is also accompanied by many questions: what is AI doing to the way we write, teach, create, judge, memorise and conduct research? What is it doing to professions, institutions and knowledge? And what does it tell us, in return, about what we thought was uniquely human? AI is no longer just a technological object: it is becoming a condition of contemporary thought itself. Through eight contributions from various disciplines, all practised at Paris 1 Panthéon-Sorbonne University, this issue of #1257 invites you to take stock of contemporary changes.

Long considered a technical field reserved for computer science labs or industrial applications, artificial intelligence has now established itself as a cross-cutting phenomenon with systemic effects on our societies. The recent emergence of generative AI (GAI) – capable of producing text, images, sounds, or videos based on a simple “prompt” – marks a decisive acceleration: software no longer merely assists human activity, it simulates human actions, automates production, and sometimes obscures information sources. Even more than previous digital revolutions, AI acts as a “cognitive technology”: it transforms the way we read, write, search, judge and create. As such, GAI is not only a subject of study: it is also becoming a condition of thought itself. It is therefore an intellectual, cultural and political issue – in a word, a civilisational one.

AI is as much a challenge as it is an opportunity for the humanities and social sciences, as indeed it is for all scientific disciplines. It calls for a renewal of practices, yet without abandoning critical standards.

The humanities and social sciences (HSS) are at the heart of the transformations underway

Economists have long used AI to analyse behaviour; historians to read ancient manuscripts; lawyers to examine the effects of law on algorithmic systems; philosophers to rethink creativity, judgement, decision-making and aesthetics, etc. AI in the humanities and social sciences (HSS) is thus much more than software that manipulates data. It changes the methods, objects and values of research itself. Each discipline shows that AI is more than, a technical breakthrough, it is first and foremost a challenge of interpretation, translation and meaning.

AI touches on the very foundations of the HSS. It transforms our relationship with sources, narratives, evidence, norms and style. It forces disciplines to review their methodologies, question their criteria for objectivity and redefines the contours of scientific authority. But it also invites us to think differently: by hybridising approaches, combining knowledge, opening archives, and exploring new formats for production and dissemination. In this respect, AI is as much a challenge as it is an opportunity for the humanities and social sciences, as indeed it is for all scientific disciplines. It calls for a renewal of practices, yet without abandoning critical standards. AI forces us to walk a ridge line, of exhibiting neither naive technophilia nor dogmatic rejection, but of adopting a posture of inquiry, questioning, and analysis, at the interface of calculation and meaning.

AI is acting as a catalyst for disciplinary restructuring

AI exposes researchers to new bodies of work, new forms of reasoning, and analytical tools from computer science. In doing so, it challenges the established boundaries between fields. Historians are collaborating with computer scientists to read in

new ways the traces of the past, economists are engaging in dialogue with computational linguists to analyse discursive flows, and artists are exploring new aesthetic spaces. AI encourages unprecedented experimentation, and sometimes unexpected hybridisations. But these technical collaborations cannot mask the epistemological tensions they give rise to: what forms of knowledge are produced by algorithmic devices? Under what conditions can we speak of interpretation by “learning” software? What collective responsibilities are emerging in these reconfigurations?

When using AI, not everything is an illusion, but nothing can be taken for granted. Discourses on efficiency, rationality and supposed “algorithmic neutrality” of AI are often driven by commercial promises or techno-political fictions. However, these promises are encountering obstacles in practice, with reinforced biases, increased opacity, new dependencies, as well as resistance to the reconfiguration of economic, political and military powers. The challenge for the humanities and social sciences is therefore not only to document AI, but to question it based on what it does to forms of knowledge, social norms and collective life. This special report aims to contribute to the critical lucidity concerning AI by offering informed, demanding and situated perspectives on a phenomenon that is constantly redefining itself.

This special issue of #1257 brings together eight articles embodying this plurality of approaches: i) an analysis of the potential uses and limitations of large language models in social sciences; ii) a philosophical examination of the disappearance of authorship in algorithmic art; iii) an exploration of disinformation in the era of GANs (*Generative Adversarial Networks*) and cognitive conflicts; iv) a geopolitical reading of how AI is altering power relationships between countries; v) a warning against the mirages of objectivity in automated recruitment; vi) a meditation on the recomposition of soundscapes situated between ecological memory and artificial simulation; vii) an overview of the uses of AI in the historical sciences; and finally, viii) a critical investigation into the transformations of journalism within an automated and monitored environment. These are all complementary perspectives, combining knowledge and awareness. Through the eight articles in this special issue, along side with other articles available online, Paris 1 Panthéon-Sorbonne University demonstrates that it is aware of the scale of the changes underway and has made AI a strategic focus for scientific, educational and institutional development. It thus joins the circle of major research universities which, at the international level, consider that the humanities and social sciences have a decisive role to play in the analysis, understanding and supervision of AI technologies. By asserting this priority, the university intends not only to support contemporary transformations, but also to shed light on them, question them and contribute to their orientation from a democratic, ethical and humanistic perspective.

The university’s broader engagements in the age of AI are also reflected in the structuring of ambitious programmes set out in the section on the University of the Future. Through these actions, the Paris 1 Panthéon-Sorbonne University is acknowledging its leading role in mobilising the humanities and social sciences around AI. It seeks to contribute fully to the construction of an informed public debate, both nationally and internationally, as well as to the consolidation of an academic space in which technical innovation cannot be dissociated from the needs of meaning and responsibility. This special issue also shows that AI is not a technological blatancy, but a major intellectual challenge at the crossroads of disciplines, societies and future democratic choices. The ambition that guides the following pages is to set the terms for a rigorous debate, rooted in the diversity of disciplinary approaches, to better think about our common future in the era of artificial intelligence. And looking beyond technical progress, the articles here seek to examine the social, economic and moral progress needs of the humanities and social sciences. ●



AI is not a technological blatancy, but a major intellectual challenge at the crossroads of disciplines, societies and future democratic choices.



Stéphane LAMASSÉ, Camille SALINESI, and Célia ZOLYNSKI



Substitution

The beginning of the contemporary era can arguably be dated to the moment when art set about erasing itself. It was as if art itself had to be destroyed, following the anonymity of artists eager to be forgotten behind their work in medieval times, and then the exaltation of artists in Romantic times. Contemporary painters buried perspective under impressions, motifs emerged in themselves, forms became sublimated gestures, and then traces and gestures were lost in colour with monochromes of black, red and blue. There followed the end of colour itself, which was replaced by holes, burns, and nothingness. In the end, neither art nor the artist had much left to say.



Philippe Huneman

is a philosopher of science and CNRS research director at the Institute for the History and Philosophy of Science and Technology (IHPST).

he fate of music has not been much better, and writers have destroyed their characters, their stories, even their descriptions and dialogues: Beckett's legless characters monologuing about their

rubbish bins are the best representation of the artist and ultimately of art, which is disappearing everywhere. It is as if the destruction of humanity, predicted some fifty years ago due to atomic bombs and failed (or successful) deterrence, had borne fruit only metonymically, with only art, an essentially or intrinsically human production, having now actually finished with itself. Or almost. Yet these artists are all avatars of characters by Kafka: here, a rodent is building a burrow that buries it; there, a singing mouse whose song constantly goes off track and turns out never to have been sung; or again, we have a champion of fasting, etc. They all struggle to reach the end of oblivion, while behind them, someone emerges to finish the job. She or he has no gender because they are not human. It is ChatGPT4 or, maybe when you are reading this text, GPT5, or 6, although I will continue to call it GPT3... Everyday language calls this "artificial intelligence". We may be moved by it, worry about it, admire its prowess or cry out that the end of the world is nigh. Sometimes we do all of these things at once. But this intelligence does something much more modest, as it indeed says of itself. It should not frighten us, as it urges us in an article entitled "Are you scared yet, humans?" which GPT3 wrote, about itself, in the New York Times (or the Guardian) in response to those who are bitter and quick to condemn the wonders of the human brain (at least for now). Artificial intelligence categorises, which, let's face it, is the basis of thought. If all apples and all oranges were the same thing, it would indeed be difficult to reason.

And above all, to take it further, if fruit and building are synonymous, just as boxwood and punishment, if a body of water and the blue of the sky can be confused, if every colour is the same as all the others: what and how should we think? The synthetic mind classifies, then recognises and discriminates. It then knows which faces inhabit your photos. Which cars drive down your street? Which shops are still open at this late hour, within an image of your neighbourhood displayed on Google Earth? And so on. But that was only GPT1. Its successors know how to perform the reverse operation.

Par Carole Raddato from FRANKFURT, Germany - Narcissus by Caravaggio, 1597 - 1599, Galleria Nazionale d'Arte Antica, CC BY-SA 2.0, <https://commons.wikimedia.org/wiki/index.php?curid=45895425>



All the images in all the museums in the world, along with all the sentences of all the spectators and critics in the world, are enough to create everything. We can stop there.

Towards a new art form

Give AI words, and it will suggest images that match. Pairs of words, groups of words, and it will do the job. You will get a series of drawings. It first takes those found on the internet, but when they do not exist, it invents them. AI creates photos of objects that are named in your prompts: “A few mismatched flowers”; “A rusty yellow scooter at the end of a dead-end street”. These are easy, of course. But let’s make them more abstract: “A jealous man sitting on a bench”; “A procession following the funeral of the great writer”; “The front of the procession”. And then there is free association, the surreal world of word combinations, the magnetic field of thoughts that Soupault and Breton spoke of: an anvil smiling at life; an approximate man; a multiplication table of torments. And there we have it. All the images are there. No more artists. Even artists’ styles, their signature, can already be adopted by simple apps on your phone.

“Warholise” a photo. “Pollockise” a few colours. “Rothkoise” your sofa. One click and you have Degottex’s splashes, Hantai’s folds, golden nebulae from Zao Wou Ki – though never painted, they could have been done by him. GPT3 will put these signatures together just as you put words and sentences together. A hybrid of Klein and Pollock, Hartung and Vieira da Silva. A Soulages of blue, a Toffoli of stripes. Give AI an image or a colour sample, and it immediately invents styles, artists and their art. All the images from all the museums in the world, all the sentences from all the spectators and critics in the world, are now enough to create everything – we can stop there. There is no more need for artists: they are all mass-produced. A total, radical erasure has been achieved. The erasure of art with the disappearance of artists; because, ultimately, what is the difference between this automatic production of paintings, sculptures, pictorial traditions, schools of visual artists, avant-gardes, permanent revolutions,... the whole jamboree of conservatives and hot-heads that each generation of art history has managed to recreate on the one hand, and the unlimited, constant production based on the incessant processing of data stored in numerous over-ventilated and overheated centres scattered around the edge of the Arctic Circle, of “content”, according to lexical usage, on the other? This term “content” is deliberately empty, neutral, and capable of naming absolutely anything. Any piece of art is content like any other: this is how art has succeeded in erasing itself. From the same source come saturated post-post-post-expressionist monochromes and pastel wallpapers where wildebeests frolic. In the same vein, there are original Brancusi-esque sculptures and advertisements for French underwear full of fresh and playful humour: all contribute to the erasure of art – and of all art. Films, according to the same principle, may be invented almost instantly by Sora – a GPT-avatar –, based on five words from the script. The graphic interfaces are so sophisticated that the actors on screen exist even more than the originals, printed on the film of yesteryear or captured by the sensor of the digital camera. “A brother-in-law discovers that his sister is leading a double life.” The film is produced without a screenwriter, without a crew, without anything; it is executed in a matter of minutes from the moment the producer types these words on a keyboard. A Marina Hands looka-

like struggles with the growing inquisition led by her husband's brother, while the latter tries to overcome his addiction to opiates, with the help of a psychotherapist played by an ageing star. There are lots of twilight scenes; a soundtrack that is hard to bear, made up of the clinking of cutlery and the demanding sound of a coffee machine in an old-fashioned café. Give Him a Beating will be a huge success. A chronicle of a group of young people in a run-down secondary school. At the end, the unpleasant teacher is hit by a train. The producer did not have to write another word; the title alone allowed GPT3 to construct all the images.

The AI web

Maliciously, using deepfake technology, the artist gave all the unpleasant characters the faces of members of the government. The worst of them all sees his collection of child pornography photos posted on social media; his wife leaves him, and he commits suicide by swallowing a bottle of bleach. The list is long, and would be tedious to enumerate. The box office now belongs to GPT3, even though for years it already belonged to some GPT that knew how to predict the public's tastes, tell which new superhero would win their favour, which teen comedy would attract housewives and their offspring. However, one musical about the misadventures of two lesbian swimmers who are injured stands out from the crowd. The songs have been played a lot, and they stick in your head. They have allowed GPT3 to specialise in different musical genres, including pop songs, orchestrations and songs with lyrics. Art has faded into the background because GPT3 produces it as spontaneously and continuously as it produces everything else: goods, advertisements, news, social networks and political opinions.

The difference between art and technique, art and industry, art and all those practices that are not art, is blurring. It's over. And even if this is not convincing, it would indeed be over because the significant difference between "art and reality" is fading ever further. For a long time now, GPT3 has no longer been content to produce films without actors or sets, making movies entirely derived from image and video databases, from the entirety of the world's existing cinema output combined with its super-powerful algorithms that no human can control. No, GPT3 knows that by creating photos and films that are indistinguishable from real things, from incarnate humans, the door is open to the complete replacement of reality – which is costly and long overrated anyway – by its own fantasies. The spectators you see fill you with doubt. Where do they come from? Do they have a home, friends, a place of birth, an origin? Or did they come into existence a few moments before you met them? Created or conceived by an algorithm which decided that, in a particular moment and place, a certain kind of person was needed, that they were the best option for fine-tuning the best of all possible worlds, as had been initially decided. You watch a science fiction film depicting this world where algorithms have taken over. You get up. You have no way of knowing who is around you, who is real and who is not, whether this cinema is in your living room, and if your living room is in a film. And this film does not film anything that really exists.



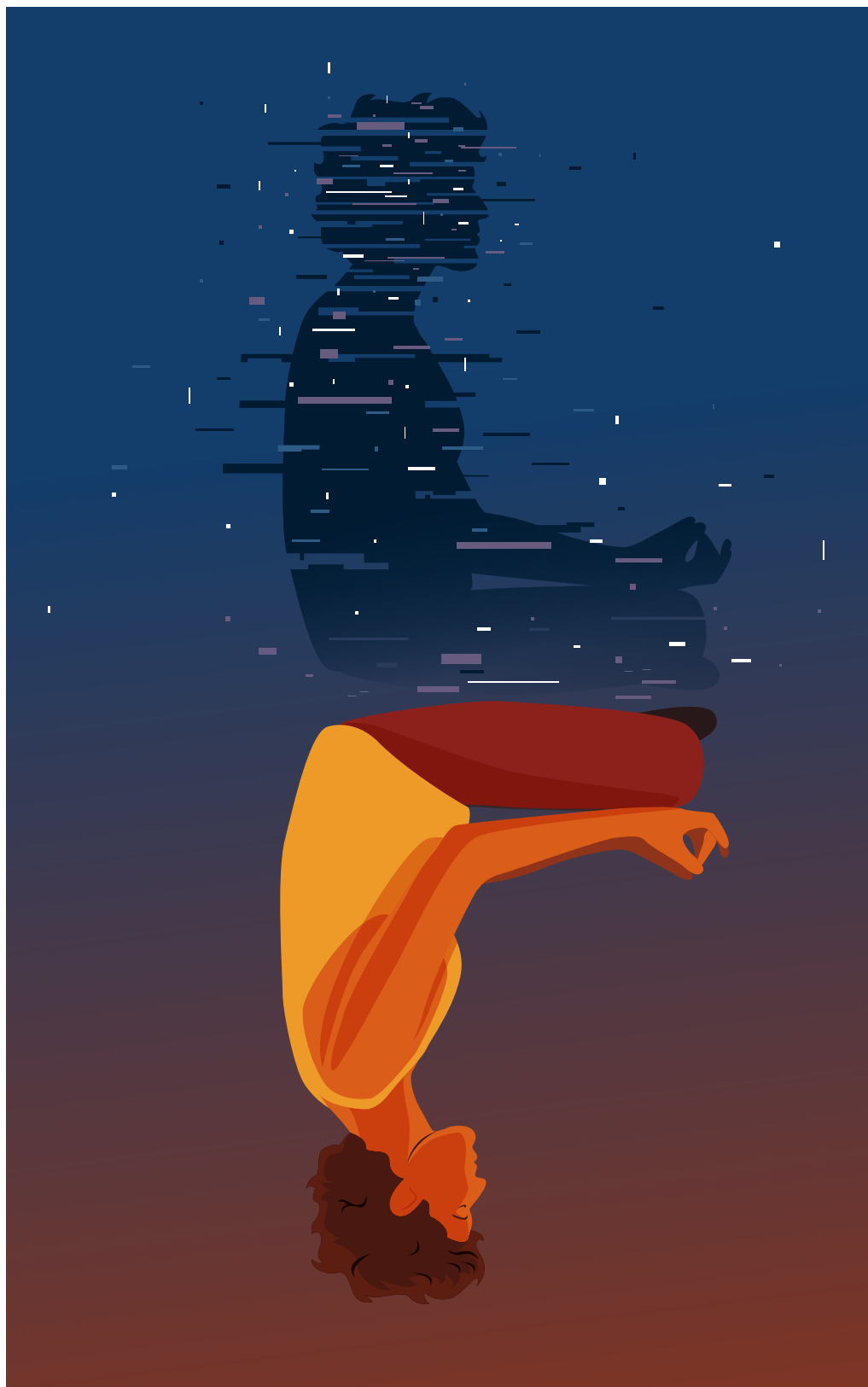
Films, according to the same principle, are invented almost instantly by GPT3, based on five words from a script.



“
**Art has faded
 away because
 GPT3 produces it
 as spontaneously
 and continuously
 as it produces
 everything
 else: goods,
 advertisements,
 news, social
 networks and
 political opinions.**”

The loop between the world, images of the world and the discourse on the world

GPT3 has commented on its own films, on its own images. It has made other films, other shots, other comments. They have escaped, and are spreading across networks, and those auxiliaries of the world which we call “digital” or the Internet. They have penetrated the interstices of the world, and mixed with the images and realities of the world – with real humans, the animals that are already there and with authentic places. They have spread, proliferated and swarmed. And, as better variants of real objects, they have supplanted them. AI has created this loop between the world, the images of the world and the discourse on the world and its images in which you – a human, like me – evolve, live, speak and die. AI creations have taken us into their loop, like an indifferent and superfluous kennel keeper would hold a sick puppy in their arms. Their words are mine, here and now. This text you are reading was probably written by GPT3, and it designed you to read it, simply, without urging you, without rushing you. You are in art, you might say to yourself, because art is everywhere. That will be little consolation. Intelligence has won. Think as much as you like, you will no longer understand it. By fading away, art has erased the world. ●



GANs and Generative AI: the Challenge of Disinformation and Cognitive Warfare

“Cognitive warfare” refers to a form of conflict in which human cognition, emotions and behaviour become the direct targets of information attacks. In this context, AI is no longer just a neutral technological tool, but a weapon for manipulating perceptions. As demonstrated by Russian information operations in the context of the war in Ukraine, the power of generative AI, when used for strategic information purposes, gives rise to a new type of conflict. We are now facing cognitive warfare. Far from being a mere metaphor, this is today an operational reality in which AI plays a central role.



Christine Dugoin-Clément

is a lecturer at the Paris
Institute of Business
Administration (IAE)
and a researcher at the
IAE Risk Chair, CREC and
CREOGN.

The war in Ukraine has brought disinformation to the forefront of public concern since 2014. This phenomenon has taken on even greater significance with Russia’s massive invasion in February 2022. Indeed, while the public has become aware that its opinions are a potential target, first with the Cambridge Analytica affair and then with attempts to interfere in various electoral processes and referendums, the invasion of Ukraine was accompanied by a particularly aggressive Russian information policy. This has placed humans – their cognitions and emotions – at the centre of the conflict.

When Russian actions, which are part of a cognitive warfare strategy, target Westerners, they attempt to weaken their support for Ukraine; when they target other areas, they attempt to strengthen support for Moscow or facilitate the penetration of a territory by actors acting as intermediaries for Russia, such as the famous private military company (PMC) Wagner, now known as Africa Corps.

Many of these operations take place in the digital realm, although more and more information operations are crossing the boundary between the real and virtual worlds by relying on actions in the physical world. One example is the Star of David tags in Paris, which have been picked up online by site networks such as *Portal Kombat*.¹ However, we have also seen an increasingly widespread use of artificial intelligence (AI) in the production and dissemination of these information operations, particularly in a cognitive warfare approach.²

¹ VIGINUM (2024). *Portal Kombat : un réseau structuré et coordonné de propagande prorusse*. Viginum.

² AUTELLET, E. (2022). *Cognitive Warfare: Contribution of the French Armed Forces Deputy Chief of Defence, NATO Collaboration Support Office*.

BÂRGĂOANU, A. and F. DURACH (2023). *Cognitive Warfare: Understanding the Threat*. Routledge *Handbook of Disinformation and National Security*, Routledge: 221-236.





Approaches targeting disinformation operations were observed during World War II, notably with Fortitude and the BlackBox operations led by Denis Sefton Delmer.



Cognitive warfare and the potential of AI

Cognitive warfare is an old concept, but like information warfare and interference operations, it has seen a resurgence of interest over the past decade. Indeed, disinformation operations were observed during World War II, notably with the *Fortitude and BlackBox* operations led by Denis Sefton Delmer.³ More recently, psychological operations, the famous Psyops, which later became MISO (Military Information Support Operations), have been in the news, not to mention the active measures widely used during the Cold War by the Soviets.⁴ *Operation Infektion*, which targeted the spread of false information about the origin of HIV, remains a model of its kind.⁵ More recently, with the resurgence of disinformation, cognitive warfare has been the subject of new research and doctrine conceptualisation projects around the world. Human cognition may therefore be a field of confrontation in the same way as other more traditional areas of conflict, such as land, sea, air, space and cyberspace. Thus, the affective, cognitive and conative aspects (the “active trinity” of human behaviour) may be the target of simultaneous or separate attacks. Similarly, different cognitive biases may not only be targeted, but also reinforced through targeting enabled by information gathering on the one hand, and the rapid production of targeted content on the other, particularly through generative AI.

In this context, the use of platforms can be particularly useful in that it allows not only the dissemination of a narrative, but also the collection of data on populations that can then be targeted according to strategies chosen. We should recall the mobilisation of platforms such as TikTok (with some 1.5 billion active users worldwide in 2024), during the 2023 German European elections⁶ and the Romanian presidential elections.⁷ However, beyond the use of social media platforms as a means of dissemination, the use of AI has taken a leap forward, both in the generation of deepfakes (which are not limited to social media or the general public) and in content creation.

AI and information operations

A *deepfake* video circulated from the outset of the invasion of Ukraine. Created by Ukrainians to raise awareness, this synthetic content showed an air strike on the city of Paris.⁸ But the video had the opposite effect: Moscow, denying the reality of the war and the information broadcast on Western channels, used it to allege that European news reports were made up of synthetic content. The number of *deepfakes* quickly

3 RICHARDS, L. (2010). The black art: British clandestine psychological warfare against the Third Reich, www.psywar.org.

4 KUX, D. (1985). “Soviet active measures and disinformation: Overview and assessment.” The US Army War College Quarterly: Parameters 15(1): 17.

5 BATES, S. (2010). “Disinforming the world: Operation INFEKTION.” The Wilson Quarterly 34(2): 13–15.

6 SCHÜLER, M., et al. (2024). Analysing TikTok’s “Others Searched For” Feature. AI Forensic.

7 BOUNEGRU, L. (2025). The great Romanian election TikTok replay, Public gata Lab.

8 Le Parisien. (2022), “« Si on tombe, vous tombez » : une vidéo choc imagine Paris bombardée par des avions russes”, 12 March 2022, <https://www.youtube.com/watch?v=aeqQ6qq9JfA>

increased, both in Ukraine and more broadly, including in the context of democratic processes. For example, deepfake audios were broadcast during the pre-election period of silence before voting in the Slovak elections,⁹ and then targeting Keir Starmer of the British Labour Party.¹⁰ However, there has been a shift in the use of mainly audio deepfakes. While the first ones sought to deceive their audience, such as the poorly made deepfake of President Zelensky broadcast at the start of the war, we quickly saw deepfakes created for humorous purposes, aimed at ridiculing an opponent or enemy – again, both in Ukraine and elsewhere. It should be noted that, in the case of this specific use, it is easier to comply with regulations requiring videos to be labelled as AI-generated: the idea is not to pass off synthetic content as authentic, but to frame perceptions or take advantage of various cognitive biases, including anchoring, memory availability and halo effects. This ensures compliance with content legislation while implementing an information strategy. This is more about diversification than replacement, as demonstrated by the broadcast of a fake news programme on France 24, which had been hacked for the occasion,¹¹ and the broadcast of pseudo-pan-African adverts supporting the new Traoré government in Burkina Faso,¹² which were in fact nothing more than deepfakes.

These uses of deepfakes must be put into perspective. Indeed, the generative adversarial networks (GANs) that produce deepfakes have made enormous progress since the first black-and-white photos produced by Goodfellow. Not only are they increasingly effective, but they can produce a wide range of results beyond just audio and video generation. GANs can also be used for image super-resolution, starting with a low-resolution image, and gradually improving the accuracy of the generator and discriminator networks. GANs can be used for image *inpainting*, for example, to restore artefacts or historical images or to remove unwanted marks from images, object recognition, and in advanced driver assistance systems (ADAS). GANs can also perform video generation and prediction, particularly to understand the movements of objects and people, or to generate animated characters, especially in video games. Furthermore, GANs can carry out text-to-image transformations, notably using *Adversarial What-Where Network* (GAWWN) technology. From a more practical perspective, they can generate spatially accurate maps of all sizes, or be deployed using different types of data, such as augmented remote sensing data. If we put this into the context of the France 24 case (mentioned above) then in the future it will be possible to influence operations that, using prior sharing, target specific segments of a population by supporting a narrative with supposedly authentic and unforgeable elements, such as maps or satellite images.

“
It should be noted that, in the case of this specific use, it is easier to comply with regulations requiring videos to be labelled as AI-generated [...]”

9 FARSKÁ, K. (2024). “Eroded resilience of democracy-Attacks on journalists as part of a political agenda and promises.”

10 BRISTOW, T. (2023). “Keir Starmer suffers UK politics’ first deepfake moment. It won’t be the last”, Politico.

11 <https://www.france24.com/fr/fr%C3%A9missions/info-ou-intox/20240216-france-24-victime-d-un-deepfake-l-intox-continue-%C3%A0-circuler-sur-le-web>

12 GLEZ, D. (2023). “Au Burkina Faso, des deepfakes au service de la transition”. Jeune Afrique.

“
**AI should not be
 made the be-all
 and end-all of
 all information
 issues: technology
 remains only a
 tool, however
 sophisticated it
 may be.**
 ”

GANs can also be used to produce sounds such as voices and music, as well as familiar noises such as engine sounds. This makes it possible to simulate sound identities that can fool the radars used in military intelligence. At the same time, GANs can also be applied to handwriting recognition. This also paves the way – in a very short time – for the imitation of handwriting, again offering new possibilities in terms of influence, disinformation and misinformation.

Finally, AIs can be used in areas such as the generation of infographics from text, as well as for website design. We have seen the particularly rapid proliferation of fake news sites generated by AIs – 900 in recent searches.¹³ To increase their effectiveness tenfold, these sites use generative AI not only to produce their content, but also to translate it into different languages. The sites mentioned above publish in some 14 different languages in order to expand their audience. Generative AI therefore allows information strategists to multiply their tools at a very low financial and human cost.

AI and disinformation as a modern pharmakon (i.e. remedy or poison)

However, AI should not be seen as the be-all and end-all of all information issues: technology remains only a tool, however sophisticated it may be. More and more AI, including open source AI, is being developed to detect *deepfakes* and even content produced by generative AI. This is particularly the case with tools such as Deepware, TrueMedia.org,¹⁴ InVID-WeVerify, and those produced by Buffalo University and Forensic++.

Yet, as in the age-old story of the policeman and the thief, every corrective measure leads to efforts to circumvent it. In cognitive and information warfare, tools are a great help. But the target and the source remain human. Therefore, if technology is to be channelled into these reprehensible uses, it is above all humans who must be made aware of the threat and adopt behaviours that protect themselves, as far as possible, from influences that could lead to decisions over which they have no control. ●

¹³ QUADRI, F. (2024). Over 900 AI-generated news websites uncovered – Report, FactCheckHub.

¹⁴ <https://www.truemedia.org/>
<https://scanner.deepware.ai/>

AI and the New Bipolar Geopolitical Landscape

AI is not creating a unified world, but a world that is once again structured around blocs. With American AI, which is focused on the market and individual freedoms (at least on the surface), and Chinese AI, which is based on state centralisation, social control and the logic of power, this bipolarisation is not only technological, but also ideological. In this context, is the geopolitical horizon of AI one of global governance or a new Cold War, in which it becomes a matter of sovereignty, domination and global confrontation are at stake?



Pierre Beckouche

Professor of Geography -
UMR Ladyss
AI Observatory at Paris
1 Panthéon-Sorbonne
University
F.R. "International College
of Territorial Sciences"

The digital revolution is rightly perceived as global. It has been one of the drivers of globalisation, and is even sometimes considered as the new spirit that would – finally! – bring together civilisations whose history has been characterised by alternations between cooperation and war. But surprisingly, digitalisation now seems to be undermining globalization, especially as AI has become a basis for innovation and efficiency in organisations and armies in recent years. AI is playing an increasingly important role in international power relations, particularly between the Western, the Russian and the Chinese worlds. We say “world” because RuNet is a key instrument for strengthening Russian authority over its regional periphery, as is Chinese AI in East Asia, and US AI, which in the West is reducing the European Union (EU) to a subordinate role.

The good old dialectic between globalisation and regionalisation...

The internationalisation of the world after World War II took two forms. We generally think mainly of globalisation, but another major trend, the increase in trade between neighbouring countries in a large region, progressed even more rapidly from the 1980s to the 2010s. Several international treaties have endorsed this regionalisation of the world. In 1992, the Maastricht Treaty transformed the European Economic Community into the European Union, with a common foreign and security policy, European citizenship and a common currency. The same trend was seen in the Americas with Mercosur in 1991, NAFTA in 1992 and the proposed Free Trade Area of the Americas in 1994. In East Asia, the ASEAN+3 agreement in 2000 was followed by the Chinese initiative that led to the signing of the Regional Comprehensive Economic Partnership in 2020. Regional integration is also taking place in Africa: ECOWAS, the East African Community, the Southern African Development Community, etc. This culminated in the signing of the African Continental Free Trade Area (AfCFTA) in 2021.

However, this dialectic has become confused. Globalisation is being challenged by the climate crisis, which promotes local or national production (“deglobalisation”), by the Covid-19 pandemic (2020-2022) and by the return of nationalism, as illustrated by the re-election of Donald



Globalisation is being challenged by the climate crisis, which promotes local or national production (“deglobalisation”), by the Covid-19 pandemic (2020-2022), by the return of nationalism, as illustrated by Donald Trump’s re-election, and, paradoxically, by the acceleration of digitalization.



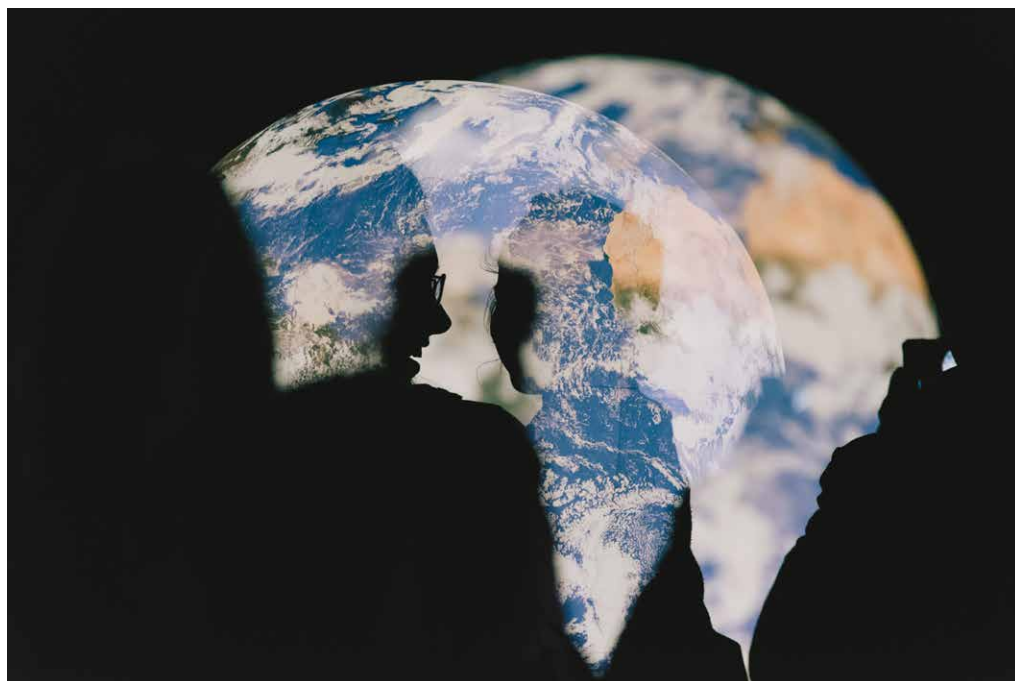
Trump. The regional organisation of the world is being challenged by the rise of the BRICS countries, and especially China, in international trade. Geopolitical tensions even point to regional disintegration, whether in East Asia with the Sino-Taiwanese conflict, in the EU with Brexit in 2016, in the wider European region, where the Arab Spring (2011) dealt a severe blow to the Union for the Mediterranean signed in 2008, and the Russian-Ukrainian war (2014 and again in 2022) that have ended the EU-Russia strategic partnership forged after the fall of the USSR [Beckouche and Richard 2023]. The emergence of AI as the next phase in digitalisation thus raises the question: are the notions of regionalisation and globalisation not in the process of being swept away by a Sino-American bipolarisation of the world?

... is being challenged by digitalisation

Digitalisation complicates the globalisation-regionalisation dialectic because several opposing visions exist. The first is that of the United States (US), which considers that the internet should be a space of free movement – under its benevolent tutelage. Its domination is systemic: its standards are imposed on microprocessors and the software that uses them, and its mastery of the ocean of data allows the US to dominate AI. One word sums this up: the “silicolonisation” of the world [Sadin 2016]. A second vision is that of Russia and China. They want state control in the face of American influence, and internally in the face of their opponents. RuNet refers to all Russian-language digital networks. Since 2013, we have seen how the Russian government has used it to annex Ukraine’s Donbass region and Crimea. Along with gas diplomacy and the Collective Security Treaty Organisation, this is Russia’s main means of controlling the peripheries of the former USSR. The 2019 law known as “Sovereign RuNet” gives the state the means to control data entering and leaving Russian territory and to isolate the Russian network, and therefore its peripheries, from the rest of the world [Limonier 2023]. Russia wants to develop its equivalent of American Big Tech and the missing components of its digital sovereignty. Ahead of many political leaders, Vladimir Putin famously proclaimed to students at Yaroslavl University in 2017 that: “Whoever becomes the leader in AI will be the master of the world”.

The Chinese case is similar, with greater technological resources (the advances in 6G mobile technology, advances in quantum communication networks and cryptography), greater industrial resources (Chinese Big Tech, and development of Chinese operating systems), as well as political resources (the Great Firewall controlling data routes and blocking IP addresses, monitoring of web activities). China accounts for 15% of total global R&D spending by digital companies, a rapidly growing proportion, while the United States accounts for 62% and Europe less than 10%, and this latter share is declining [Lhuillery 2021].

The third vision is that of the EU. While many IT innovations originated in Europe, Europeans are paying the price for their financial culture, which is less open to risk-taking, and the lack of a common strategy: they have not managed to achieve in the digital sector what they did in aero-



nautics with Airbus fifty years ago. And when action does manage to be confederal (at the EU level), it is geared towards protecting users rather than entrepreneurial innovation, such as: i) the General Data Protection Regulation (GDPR, 2018); ii) the Data Act on the data we generate with our connected objects (2023); iii) the Digital Markets Act to combat abuse of dominant position by Big Tech; iv) the Digital Services Act to combat disinformation or racism (2023); and v) the Artificial Intelligence Act “centred on people, [to be] ethical, sustainable and inclusive” (2024). But in semiconductors, Europe accounts for only 8% of world production. Launched in 2020, its Gaïa-X cloud project involves companies such as Amazon, Microsoft, Google (which combined account for 69% of the European cloud market), Alibaba and Huawei. In short, we have the right rules, but the Chinese and Americans have Big Tech. And they are also fighting on the regulatory front, particularly the United States. First, the US is competing with European regulation: since 2018, the Cloud Act (Clarifying Lawful Overseas Use of Data) has allowed the US government to request emails, documents and personal data from any American company, even if its servers are located in Europe, in contravention of Article 48 of the GDPR. Secondly, to criticise what he sees as any “excess” of public regulation, President Trump, “to liberate innovation” as he put it, repealed Executive Order 14110 signed by Joe Biden in 2023 to protect privacy from AI at the end of January 2025.

AI: towards a new bipolar geopolitics

The realities of AI are shaping a new geopolitics. Russia’s AI plan suffers from structural limitations, both in semiconductors and algorithms. The war in Ukraine has limited the civilian deployment of Russian AI in

“
China accounts for 15% of total global R&D spending by digital companies, a rapidly growing proportion, while the United States accounts for 62% and Europe less than 10%, and its share is declining.
”

The EU trains more AI engineers than the United States and far more than China, but many end up on the US labour market. Per capita, a country like France invests three times less in AI than the United States.

favour of military priorities, and even then, with limited success except in disinformation. Since 2022, the brain drain of Russian AI talent has been accelerating, including to China [Nocetti 2025].

European AI has still not really taken off [Babinet 2025]. It suffers from a lack of financial integration and stronger partnerships in R&D. The price of Brexit is high because the United Kingdom was, far ahead of Germany, the leading country in European AI alongside France. Moreover, the European AI Act would cost the European economy €31 billion and reduce investment in AI by 20% [Mueller 2021]. It could penalise European start-ups because they will have to assess whether the AI they develop could pose a risk to humans, and if so, at what level of risk. The EU trains more AI engineers than the United States and far more than China, but many end up in the US labour market. Per capita, a country like France invests three times less in AI than the United States [Aghion 2024].

With Russia overtaken and the EU on the verge of vassalage, the United States may enter into an explicit power struggle with China. In 2017 the US banned the sale of semiconductors and operating systems to Chinese companies by US firms, and in 2020 extended the ban to firms from all countries. It excluded Huawei from its smartphone market on the grounds of risks of personal data collection. Since 2024, US funds have been prohibited from investing in Chinese companies specialising in semiconductors, quantum computing and all other components of AI. As for US Big Tech companies, they are increasingly convinced that the issue at stake is not merely regulation versus corporate profits, but rather American AI aligned with certain liberties versus Chinese AI aligned with government strategy [Alexandre 2023, Beckouche 2025].

China's investments in AI are immense, from GPUs to generative AI models, as demonstrated by the spectacular success of DeepSeek in early 2025. Since 2019, there have been more Chinese scientific articles on AI than American ones. Between 2013 and 2023, private investment in AI will have amounted to \$336 billion in the United States, \$104 billion in China and \$50 billion in Western Europe [Perrault and Clark 2024]. China is more credible than Europe when it comes to launching a satellite fleet capable of competing that of the Americans whose importance was demonstrated when Elon Musk provided Ukraine with free high-speed internet via his Starlink satellites and then deactivated the service within a 100 km radius of the Crimean coast to prevent an attack on a Russian naval base.

These geopolitical realities far outweigh the hopes of a globalised AI supporting development and inclusion. In 2024, Bill Gates predicted that the adoption of AI would be widespread in the global North by 2025 and in the global South by 2026. At present, Africa, which accounts for 17% of the world's population, produces less than 1% of its AI.

Sino-American digital domination could recreate a world order reminiscent of the Cold War, with two superpowers controlling the developing world, apart from India, even though AI there is largely funded by US Big

Tech. The UN could once again face difficult times. In July 2023, its Secretary-General promised to ban AI in weapons of war. But at the same time, the US Air Force announced that it was developing fleets of drones and fighter jets piloted by AI, which would give them ultra-responsiveness on the battlefield. Since 2020, human pilots have been beaten by AI pilots in simulated air combat, and since 2022, the Chinese army has been publishing similar results.

US Big Tech’s announcements regarding data centres for generative AI run into the several hundreds of billions of dollars, compared to the €109 billion announced by France at its AI Summit in February 2025. Moreover, these American projects are increasingly located in the United States to protect data. It is not impossible that access to data, computational methods, content produced and the societal impact of AI will diverge between the United States and China. Although the world is more interconnected than ever, AI could thus usher in unprecedented geoeconomic and geopolitical conflicts. ●

“
China is more credible than Europe when it comes to launching a satellite fleet that can compete with the Americans, and to its assertive strategy to position itself as the leader of the digital world.
”

RÉFÉRENCES

AGHION, Ph. (dir.), 2024. “IA : notre ambition pour la France”, The Artificial Intelligence Commission’s Report to the French Prime Minister.

ALEXANDRE, O., 2023. *La Tech. Quand la Silicon Valley refait le monde*. Paris, Le Seuil.

BABINET, G., 2025. Série “L’intelligence artificielle, une arme géopolitique”. Paris, France Culture.

BECKOUCHE, P., RICHARD, Y. (Eds.), 2023. *Regionalization of the world. Comparing regional integrations*. London, ISTE-Wiley.

BECKOUCHE, P., 2025. *Homo Externatus, l’homme procédural. La victoire de l’arithmos sur le logos*. Paris, Classiques Garnier.

LHUILLERY, S., et al., 2021. “La R&D des groupes français et le CIR”, Neoma Report for the National Commission for the Evaluation of Innovation Policies. Paris, France Stratégie.

LIMONIER, K., 2023. “Le Runet, région du cyberspace ?”, n P. Beckouche and Y. RICHARD, 2023.

MUELLER, B., 2021. “How Much Will the Artificial Intelligence Act Cost Europe?”, Washington Center for Data Innovation.

NOCETTI, J., 2025. Série “L’intelligence artificielle, une arme géopolitique”. Paris, France Culture.

PERRAULT, R., Clark, J. (Eds.), 2024. “Artificial Intelligence Index Report 2024”, Institute for Human-Centered Artificial Intelligence, Stanford University.

SADIN, É., 2016. *La silicolonisation du monde. L’irrésistible ascension du libéralisme numérique*. Paris, L’Echappée.

The Use of AI in Recruitment and the Risks of Mirages

The deployment of AI algorithms in the recruitment process is a prime example of both the opportunities offered and the risks involved when these systems are used in operations that may affect the professional futures of the individuals who use them or who are exposed to them.



In 2024, around 40% of French companies with more than 1,000 employees had integrated at least one AI solution into their recruitment process...and this figure will undoubtedly have increased by the time this article is published.¹ The trend is the same among candidates: according to a recent survey by France Travail, 83% of candidates under the age of 25 use AI in their application strategy. The recruitment process is particularly interesting as a field of study for the deployment of AI in the workplace for several reasons:

- It is a high-stakes operation for the candidate, who is gambling a part of their professional future, but also for the recruiter, given that more than 40% of industrial companies anticipate recruitment difficulties in early 2025, according to France's National Institute of Statistics and Economic Studies (INSEE).
- It is a highly regulated process, particularly due to the risks of arbitrariness and discrimination against candidates.
- Finally, this is one of the areas in which AI solutions with a direct impact on individuals' lives are developing most rapidly. The highly competitive "augmented recruitment" market was worth \$3 billion in 2024, with an expected annual growth rate of nearly 10% until 2031.

Faced with this rapid development, research is struggling to keep up and there is still little empirical work on the objective performance of these systems and on the behaviour of users and those exposed to them.

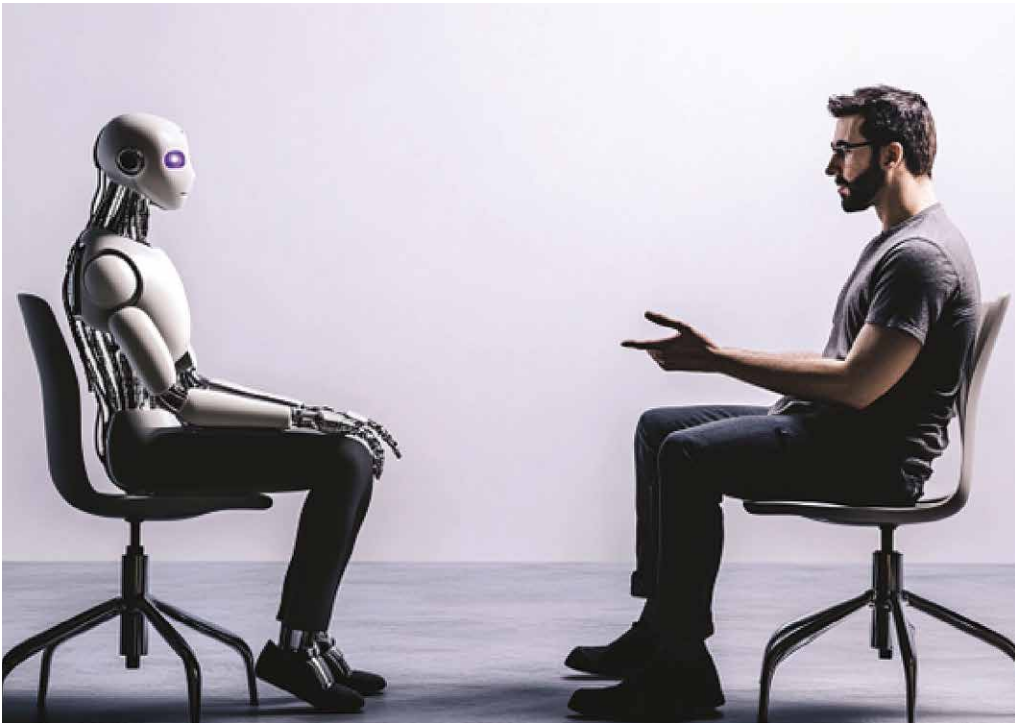
This article aims to show that AI-based recruitment tools offer undeniable advantages, which are one of the reasons for their rapid adoption. However, they also raise many technical, ethical and regulatory questions that have not yet been sufficiently studied.

Taking the example of discrimination in hiring, we will see in particular that these tools do not always help to reduce discrimination and that they can even create new biases in judgement, such as an exaggerated belief in the reliability of the answers provided by algorithmic systems.

Alain Lacroux

professor of management sciences at the Sorbonne School of Management (SSM - UFR 06) and member of the Interdisciplinary Research Centre in Management Sciences (PRISM - UR 4101).

¹ *Actua*, 2 January 2025: "L'IA, alliée stratégique des professionnels du recrutement : une étude d'ICIMS", available online.



The promising prospects of “augmented” recruitment

In practice, all stages of recruitment are impacted by the introduction of AI algorithms, from job profiling to the final selection of successful candidates, as shown in the following diagram:

Recruitment phases	AI-integrated solutions	Promises/expected benefits
Sourcing (candidate search) ▼	➤ Web scraping / Building prospect databases	➤ Diversification of applications & proactivity
Drafting of job offers ▼	➤ Semantic tools for evaluating job offers Use of generative AI based on language models	➤ Unbiased job postings
Analysis of applications ▼	➤ Automated file analysis (information extraction) Natural language processing (NLP) for cover letters	➤ Automated analysis of applications (criteria based solely on job profile)
Candidate Qualification (questions/answers) ▼	➤ Chatbots (conversational agents) ➤ Recorded interviews Automated or generative AI-assisted responses	➤ Fast and objective prequalification. Customization possible
Cognitive and personality tests ▼	➤ AI-assisted cognitive tests AI-assisted exercises, games, and personality tests	➤ Objective and unbiased ranking based on skills and personality
Interviews ▼	➤ Automated analysis of recorded interviews. Analysis of speech, prosody, vocabulary, etc.	➤ Analysis of speech and nonverbal cues (sincerity?)
Matching and ranking ▼	➤ Classification and prediction/explainability algorithms	➤ More objective and effective selection (multiple criteria)
Decision support	➤ Presentation of ranked skill profiles List of candidates with multi-criteria evaluations	➤ Assistance in making choices free from discriminatory bias

According to Lacroux & Martin (*Management et Avenir journal*, no. 122, 2021)



Faced with this rapid evolution, research is struggling to keep up and there is still little empirical work on the objective performance of these systems.



The stages of AI “augmented” recruitment

At first glance, these promises are appealing, and there is no doubt that AI algorithms can speed up, automate and reduce the costs of a number of routine operations, such as extracting information from CVs or generating automated responses to applications. It should be noted that the arguments are not only technical and economic – the promise of efficiency and profitability – but also ethical, through the promise of objectivity and unbiased recruitment.

The numerous individual psychosocial biases that impact recruiters’ judgement (the halo effect, confirmation bias and stereotyping) could indeed be limited by the automated pre-selection process and the presentation of candidate profiles chosen solely on the basis of their qualities. It should be noted that predictive recruitment tools are always presented as decision-making aids, and never as entirely autonomous solutions.

Despite this optimistic view – widely promoted in professional circles by advocates of augmented recruitment software solutions – these tools actually pose multiple problems, and decision-making biases remain. A few examples are given here by using a typology that distinguishes three types of bias related to computer systems: i) pre-existing biases, which affect the data on which the tools rely to make predictions; ii) technical biases caused by the functioning of the system (algorithmic biases); and iii) finally, emerging biases caused by the user of the system.²

Pre-existing biases: “garbage in, garbage out”

This well-known adage among computer scientists, applied to augmented recruitment, simply means that the quality of the results obtained depends on the quality and quantity of the data used to train the predictive models.

In the field of recruitment, training data mainly consists of data available on applicants, whose characteristics are cross-referenced with those of employees who were actually recruited (the “right candidates”) to isolate the best predictors of success in the job. Two mechanisms may lead to discrimination at this stage.

- If the “good candidates” used as a reference in the algorithm reflect a bias (e.g. assertiveness and leadership skills associated with masculine traits), these biases will be reproduced in the recruitment recommendations.
- If predictive recruitment tools draw probabilistic conclusions from a biased sample of the population (e.g. a sample in which women are under-represented), any recruitment decision based on these conclusions will systematically disadvantage candidates who are under-represented in the training data.

It should be noted that biases related to training data typically clone the existing employee population by recommending candidates similar to those already holding positions.

2 FRIEDMAN, B., & NISSEBAUM, H. (1996). “Bias in Computer Systems,” *ACM Trans. Inf. Syst.*, Vol. 14 (3).

Another more unexpected pre-existing bias concerns the gender of the algorithm creators (the programmers).³ The computer programming environment is indeed unique: male graduates are vastly overrepresented. Considering that the choice of criteria can be decisive in calculating a recommendation score, placing excessive weight on the results of mathematical logic tests when recruiting a manager may exaggerate the number of male candidates with a background in mathematics (who are more numerous in engineering schools), even though the link between managerial skills and mathematical ability has never been established.

Algorithmic bias: the black box problem... but not only that

The first types of algorithmic bias are related to technical failures: for example, algorithms struggle to recognise certain facial expressions or to decipher the meaning of certain expressions in natural language. This type of technical error is a potential vector for discrimination. For example, foreign accents are more difficult to “understand” during verbal exchanges with a recruitment chatbot.

As for the “black box effect”, it is now well documented: certain algorithms used in predictive models for augmented recruitment (e.g. neural networks or random forests) are fundamentally opaque, in the sense that it is not possible to detail the mechanisms linking their input and output variables. This black box effect is particularly problematic in the field of recruitment, as this operation is highly regulated by law, with several provisions aimed at protecting candidates.

The General Data Protection Regulation (GDPR) requires, in particular, that any information or communication relating to the processing of personal data be transparent and understandable, and that any person has the right not to be subject to a fully automated decision that has legal effect or significantly affects them. The French Data Protection Authority (CNIL) has also established the principle of data minimisation, which requires that the data collected on a candidate be adequate, relevant and limited to what is necessary in relation to the defined purpose. European regulations on AI have recently supplemented these provisions by creating a risk-based typology of AI systems. Algorithmic recruitment systems are classified as “high-risk AI”: certain operations are explicitly prohibited (detection of emotions or analysis of facial expressions), while the marketing of algorithmic systems is regulated. These requirements have long been considered incompatible with the “black box effect”, but it should be noted that spectacular technical progress is now being made in the field of explainable AI, whose tools are designed to provide users with assistance in interpreting results. Indeed, it is possible to specify which variables collected played the most important explanatory role in the model (a general explanation), or even to provide an explanation to a candidate by highlighting the variables that had a favourable or unfavourable influence on the decision concerning them (known as a “local” explanation).



It should be noted that predictive recruitment tools are always presented as decision-making aids, and never as entirely autonomous solutions.



³ See, for example, NADEEM, A., MARJANOVIC, O., & ABEDIN, B. (2022). “Gender bias in AI-based decision-making systems: a systematic literature review”, *Australasian Journal of Information Systems*, 26.

The ‘black box effect’ is particularly problematic in the field of recruitment, as this process is highly regulated by law, with several provisions aimed at protecting candidates.

Emerging biases: the problem of machine heuristics

Machine heuristics, sometimes referred to as “automation bias,” arise when recruiters give decisive weight to information from an algorithmic system (for example, the candidate ranked first on a recommendation list is highly likely to be chosen due to confidence in the ranking algorithm and the difficulty of the task). According to some research, trust in algorithms in ambiguous decision-making situations (which is typically the case in recruitment) is higher than trust in humans, except among experienced professionals. However, our own work has highlighted a paradox: even though recruiters say they trust an algorithmic system less than a human expert, they are more likely to follow algorithmic advice in complex situations (in this case, ranking CVs in response to a job offer).⁴

To conclude: an impossible trade-off?

The biases associated with the training data used in predictive models are a prime example of the difficult trade-off between accuracy and legality in the field of recruitment. This can be summarised as follows: the more algorithms are trained on a variety of personal variables, the more effective they are and the less ethical, legitimate or legal they are. On the other hand, the more regulatory requirements on personal data protection are respected, the less usable the available data is, which impacts the performance of the algorithms... This trade-off between fairness and accuracy is well known to machine-learning specialists, but in recruitment, it is near-impossible to resolve. ●

⁴ See LACROUX A. & MARTIN-LACROUX C. (2022). “Should I trust Artificial Intelligence to recruit? Recruiters’ perceptions and behaviours when faced with algorithm-based recommendation systems during resume screening”, *Frontiers in Psychology*, doi: 10.3389/fpsyg.2022.895997

Soundscapes and AI: Ecological Issues

Listening, recording and then composing based on soundscapes encourages a sensitive and ecological approach to the world of sound. This process of fragmentation and recombination echoes the dynamics of artificial intelligence. These practices are part reflection on memory and creative approaches, and they can be a means of transmitting and preserving our sound heritage.

The study of soundscapes and composition

The emergence of the study of soundscapes, led by R. Murray Schafer in the late 1960s, and the famous World Soundscape Project (WSP) programme of study of sound ecology and soundscapes, gave rise to practical analyses and research in the fields of listening to the sound environment and ear training, as well as sound creation. In his book *The Tuning of the World* (1977), Schafer emphasises ear training and conscious listening to the soundscape. With this approach, the whole body is transformed into an auditory receiver and becomes the very embodiment of the soundscape. For Schafer, soundscape studies lie between the arts, sciences and social sciences. This multidisciplinary approach allows us to understand the impact of sound on human life and vice versa. Thus, scientific research on psychoacoustics and acoustics, combined with a creative approach derived from music and sound art, feeds into reflections on soundscape studies. This interdisciplinary and collaborative approach has emerged as a new discipline under the term ‘sound design’: ‘A new interdiscipline [...] attempts to discover principles by which the aesthetic quality of the acoustic SOUNDSCAPE may be improved’.¹

Sound ecology, as a method of field research, requires precise and in-depth observation of sound environments. In this context, recording is a key tool.

*[Let us] regard the soundscape of the world as a huge musical composition, unfolding around us ceaselessly. We are simultaneously its audience, its performers and its composers. Which sounds do we want to preserve, encourage, multiply? When we know this, the boring or destructive sounds will become conspicuous enough and we will know why we must eliminate them. Only a total appreciation of the acoustic environment can give us the resources for improving the orchestration of the soundscape.*²

Schafer considers the sound recorder to be a complementary device for the ear. Not only does it make it possible to isolate and study a sound in “high fidelity”, but it also allows us to preserve the sound mark of

Azadeh Nilchiani

is an associate researcher at the ACTE Institute (Arts, Creations, Theories and Aesthetics) and a post-doctoral researcher for the DEMARTS project (Creation, Democracy and Digital media), under the supervision of Sandra Laugier (full professor, ISJPS).

¹ Raymond Murray Schafer, *The Soundscape: our sonic environment and the tuning of the world*, Rochester, Vermont, Destiny Books, 1977, p.271. https://monoskop.org/images/d/d4/Schafer_R_Murray_The_Soundscape_Our_Sonic_Environment_and_the_Tuning_of_the_World_1994.pdf

² *Ibid*, p.205.



“
This
multidisciplinary
approach allows us
to understand the
impact of sound on
human life and vice
versa.”

our time: that of machines or natural sounds that are disappearing or changing day by day.

The Canadian composer Hildegard Westerkamp was part of *the* World Soundscape Project. She was responsible for listening to and describing the field recordings made by other members of the team. In early 1977, Westerkamp founded a cooperative radio station and organised the weekly Soundwalking programme, which was the origin of the Soundwalk movement: “to encourage the participant to listen discriminatively, and moreover, to make critical judgments about the sounds heard and their contribution to the balance or imbalance of the sonic environment”.³

Mention must be made of the pioneering work of composer Pierre Mariétan on sound ecology, listening and the notion of the «sound environment», which he first evoked in 1969. Taking into account the «existing sound» in everyday life and applying it to the analysis of sound creation is a necessary prerequisite. On a fundamental question concerning the impact of sound technologies, Mariétan suggests that their influence can be positive, provided that they are integrated into an ecological and social approach. “The requirement for sound quality in the environment in which we live becomes necessary when it is possible to act on it with an awareness of the sounds we produce”.⁴

Since the late 1960s, Bernie Krause, a pioneer in soundscape ecology, has contributed greatly to the recording, preservation and appreciation of natural soundscapes through his Wild Sanctuary Project. He has succeeded in collecting and preserving traces of many natural soundscapes that have now disappeared, including more than 2,000 different types of biospheres, both marine and terrestrial.

The World Forum for Acoustic Ecology (WFAE), established in 1993, brings together groups and individuals who produce studies on soundscapes. Jim Cummings and Steven M. Miller (2007) discuss the role of sound art as an effective means of demonstrating changes in our sound environment and preserving our world, but in a different way from environmental activism.⁵ Today, advances in the practice of collecting, field recording, studying and analysing soundscapes have not only promoted research and initiatives addressing ecological issues, but have also given rise to artistic creations, often in collaboration with scientists and ecologists. These collective approaches have made it possible to address this subject in a way that is both sensitive and sensory.

3 Barry Truax, *Handbook for Acoustic Ecology*, Second Edition, 1999, Originally published by the World Soundscape Project, Simon Fraser University, and ARC Publications, 1978. Accessed [5 February 2025]. <https://www.sfu.ca/sonic-studio-webdav/handbook/Soundwalk.html>

4 Pierre Mariétan, “Le son partout, tout le temps : où est la musique ? Technologies du sonore : un bien pour un mal ?”, in *Sonorités Ecologie sonore Technologies Musiques*, N.6, Edited by Pierre Mariétan and Roberto Barbanti, Nîmes, Champ Sociale, 2011, p.98.

5 Jim Cummings and Steven M. Miller, “Report from the Chair,” *Soundscape, The Journal of Acoustic Ecology, Art, Science, Environment, Activism*, volume 7 number 1 | fall/winter 2007.



Technology, AI and sound creation

Telecommunications, new technologies, digital technology, particularly sound processing and broadcasting devices, and the integration of artificial intelligence into the creative process have greatly contributed to the creation of immersive, multisensory and sometimes interactive installations and environments. Listening is emerging as a powerful means of refining and deepening our ecological awareness. Listening to our environment can lead us to make recordings that capture the sounds that have caught our attention, thus preserving the sound imprint of a specific moment and place. These recordings can become the raw material for a sound creation. This process, known as “soundscape composition”, is a way of reflecting our environment. This compositional approach is based on recorded fragments of the sound environment, which are recombined in a new and distinct way from the original environment, while retaining part of their sonic identity. This process involves fragmentation, reinterpretation and rearrangement. Although this compositional practice has existed since it became technically possible to record sound – and well before the era of AI – the logic of recomposition based on fragments can be considered to be linked to the learning and reconfiguration processes of AI.

In “The Aesthetics of Fragments,” Lev Manovich evokes human cultural history and highlights a massive forgetting as well as limited preservation, brutally reducing history to a few emblematic fragments.⁶

⁶ Lev Manovich and Emanuele Arielli. *Artificial Aesthetics: Generative AI, Art and Visual Media*, 2024. <https://manovich.net/index.php/projects/artificial-aesthetics>.

“**Taking into account the “existing soundscape” in everyday life and applying it to the analysis of sound creation is a necessary prerequisite.**”

“
Listening to our
environment can
lead us to make
recordings that
capture the sounds
that have caught
our attention, thus
preserving the sound
imprint of a specific
moment and place.”

He thus emphasises one of the most interesting aspects of AI aesthetics, which lies in its ability to offer a less rigid representation of memory than that of printed (or, by extension, physically recorded) documents and resources. Even if AI does not eliminate “forgetting”, but is “slightly” less brutal in reducing representation, it permits the resurgence of cultural fragments that would otherwise have remained hidden. This is an approach in which collective memory becomes more malleable, more accessible and less subject to the constraints of time, physical media and geographical location.

Sound capture and recombination in soundscape-based composition offer a sensitive and ecological approach to our sound environment, in a similar process of rearranging cultural fragments. For its part, AI enables the reinterpretation and reappearance of memory traces. It may be asked if this is a process in which fragmentation and recombination become a means of ensuring the continuity of transmission and raising awareness of our environmental sound heritage.

The sound work and fragments of the landscape

The sound installation *Nature Manifesto*, presented between November and December 2024 at the “Biodiversity: what culture for what future?” Forum at the Centre Pompidou in Paris, was a creation by the Icelandic singer and musician Björk and artist Aleph Molinari. Lasting 3 minutes and 40 seconds and broadcast on 70 speakers, the work blended natural soundscapes with the cries of endangered animals. Björk’s voice recited a reimagining of the text of the Cornucopia manifesto, which the authors wanted to rethink in a science fictional way.

[T]his immersive sound piece gives endangered and extinct animals a voice by merging their sounds with our words. We wanted to share their presence in an architecture representing the industrial age, far away from nature. In the veins of the escalator of the museum, known as the “caterpillar,” we wanted to remind citizens of the raw vitality of endangered creatures.⁷

The work was designed with AFTER (Audio Features Transfer and Exploration in Real-time), developed by the Institute for Research and Coordination in Acoustics/Music (IRCAM), Paris. “AFTER is a diffusion-based generative model that creates new sounds by mixing two sources: one audio stream to set the style or timbre, and another input (either audio or MIDI) to shape the structure over time”.⁸ By gradually adding noise to a signal, this approach produces high-quality sounds by capturing complex characteristics of the audio data. *Nature Manifesto* uses AFTER to analyse, fragment and reassemble natural and human sounds into a hybrid recomposition. In Björk’s words:

I tried to approach the creatures on an equal footing sonically, as collaborators. I spent weeks listening to recordings made over the last hundred years by

⁷ Björk and Aleph, *Nature manifesto* –Björk.fr. <https://www.bjork.fr/nature-manifesto-bjork-aleph>. Accessed [15 February 2025].

⁸ IRCAM. “AFTER.” *IRCAM Forum*, <https://forum.ircam.fr/projects/detail/after/>. Accessed [10 February 2025].

*the BBC and nature enthusiasts – the David Attenboroughs of humanity – who waited for hours with a microphone in the wild for their protagonists.*⁹

The soundscape thus created is the result of deconstruction and recombination according to an algorithmic logic, a reinterpretation of reality through the manipulation of sound fragments. The work was conceived and designed with an approach that considers the connection between the inhabitants of nature, animals, insects, their environment and us humans. “What is important in Nature Manifesto is to bring this idea of interconnection and the urgency of saving our ecosystem and its biodiversity to a place in the heart of an urban environment”.¹⁰

Changes in a natural or urban environment have an impact on the sound environment. While some sounds appear, others disappear. Sound creations based on these sound spaces, which change over time, are both a direct witness to and a reflection of some of these changes through their imprint. This not only allows us to keep track of the transformation of the soundscape, but also to gather information about the sound space produced collectively, the influence of human activities on the environment, and to provide us with important markers of the history of places and their inhabitants. Artificial intelligence can open up new creative perspectives by allowing these sound fragments to be recomposed to create soundscapes where memory and imagination intertwine to offer a conscious and sensitive reinterpretation of our environment. ●



Even if AI does not eliminate “forgetting”, and is “slightly” less brutal in reducing representation, it permits the resurgence of cultural fragments that would otherwise have remained hidden.



⁹ Björk, in Björk, *Aleph Molinari and Ircam: Nature Manifesto*, by Thom Waite published in Dazed, 3 December 2024, <https://www.bjork.fr/Dazed>. Accessed [12 February 2025].

¹⁰ Aleph Molinari. *Ibid.*

AI, Historical Research and the Maturing Relationship between History and Computer Sciences

Artificial intelligence (AI) is profoundly transforming the way history is studied, analysed and disseminated. This recent development has deep roots, as shown by the RESEDA project launched in 1976. This pioneering project aimed to create a prosopographic database of the Middle Ages, using an AI approach influenced by expert systems. It was designed to read historians' notes and integrate them into a database. Today, AI has evolved considerably thanks to the development of artificial neural networks, to such an extent that these have become synonymous with artificial intelligence itself. What was once just one branch of AI has become its dominant symbol, eclipsing other approaches such as expert systems and formal logic, which once dominated the field.



Stéphane Lamassé

is a lecturer in history, civilisation, archaeology and the arts of the ancient and medieval worlds and a member of the Laboratoire de médiévistique occidentale de Paris (LAMOP - UMR 8589).

Léo Dumont

is a professor of contemporary history and digital humanities at the Sorbonne School of History.

Advances in automatic pattern recognition techniques, not only for printed text but also for handwriting, offer new perspectives in terms of the scale of corpuses. However, beyond its documentation aspects, the use of AI by historians, although still in its infancy, raises new questions about the use of deep learning.

The consequences for documentation production and processing are numerous. That is why we address the issue of changes in the work of historians here, by focusing on textual documentation. This evolution is changing the research and analysis practices of historians, opening up new methodological perspectives that will only be touched upon briefly.

AI is redrawing the boundaries of the historical textual corpuses: rediscovering ancient texts

AI is revolutionising the restoration and analysis of ancient texts, providing historians and philologists with tools of unprecedented precision. These technological advances are helping to preserve humanity's written heritage, but also to interpret it with new depth. The project led by Melissa Terras, as early as 2006, to decipher the tablets from the Roman camp at Vindolanda demonstrated the potential of AI in this field. AI technology proposed interpretations of the missing fragments, which researchers could then evaluate and select. The *Pythia* project, launched in 2017 by the University of Oxford, perfectly illustrates this development. Using multi-layered neural networks, the *Pythia* model is able to recover missing characters from damaged texts. Its successor, *Ithaca*, developed in 2022, takes these capabilities even further.



These innovations are part of a promising line of research. The growing interest in these technologies sometimes attracts major investment. The *Vesuvius Challenge*, launched in 2019, is one such example. With a budget of \$1.5 million from Silicon Valley investors, this project aims to decipher the *Herculaneum papyri* without opening them, using computer vision algorithms. The methods implemented help philologists, palaeographers and codicologists, but also allow a wider group of historians to access these texts, and hypotheses of reconstructions, enabling them to consult fragile documents without risking damage. This is one of the key points of the *Vesuvius* project. Based on our current knowledge, these techniques have been used most extensively for ancient periods and have had an impact on our understanding of those eras. One aspect of *The Electronic Babylonian Literature Project* aims to determine the likely geographical areas for Akkadian cuneiform tablets produced in Mesopotamia.

“**Advances in automatic pattern recognition techniques, not only for printed text but also for handwriting, offer new perspectives in terms of the scale of corpuses.**”

The transformations of HTR: algorithmic mediation for accessing and manipulating text

Handwritten Text Recognition (HTR) is a technology designed to identify signs such as letters, numbers, symbols or glyphs in digital images. Unlike optical character recognition (OCR), which processes printed characters in isolation, HTR often has to deal with handwritten cursive and highly varied scripts, which has long made the task more complex. Yet, the re-emergence of neural network-based algorithms over the past fifteen years has enabled a remarkable level of recognition to be achieved. This has never been seen before.

“
Using multi-layer
neural networks, the
Pythia model is able
to recover missing
characters from
damaged texts.
”

A major milestone was reached in 2009, when the first fully neural systems outperformed traditional statistical methods in international evaluations. In a competition to recognise Arabic handwriting, a neural model achieved an accuracy of 91.4%, far surpassing all other solutions, the best of which achieved lower results (87.2%). This was despite the fact that its creators did not understand the Arabic language.¹

On 23 June 2022, a symposium entitled *Ancient Documents and Automatic Handwriting Recognition* was organised by France's École nationale des Chartes. One of the projects presented at the conference was *eScriptorium*. Launched in 2018 at the École pratique des hautes études, it aimed to develop an integrated platform dedicated to the transcription, annotation and publication of historical documents.²

Between these two dates, the number of projects grew rapidly, despite the high costs associated with machine learning. Furthermore, this phase was not as simple as it might seem: the results obtained often required manual post-processing before they could be used. One such major project was launched at Paris 1 Panthéon-Sorbonne University: the Notre-Dame de Paris and its cloister project (*Notre-Dame de Paris et son cloître*) funded by the ANR e-NDP, and conducted between 2021 and 2025. One of the project's objectives was to set up a collaborative digital edition of 26 medieval registers (1326-1503) from the cathedral chapter, a community of 51 canons exempt from any royal or municipal authority, up until the French Revolution in 1789.

Despite these challenges, advances in HTR have profoundly transformed the approach of researchers in the analysis of manuscript texts. It is now possible to extract useful information even from documents whose transcription is not perfect, and the range of digitised documents available has diversified. Today, this technology is applied to a variety of media as diverse as statistical tables, medieval books of hours, contemporary correspondence, and even comic strips.

This transformation in access to texts and documentary information inevitably raises questions about the work of professional historians today, as these are the media on which their thinking is based, constituting one of the methodological foundations of the discipline. More generally, AI algorithms not only enable reading, but also the extraction of various pieces of information that can challenge our view of documents by drawing greater attention to their actual production. Natural language processing (NLP) techniques are becoming easy to use, especially as large language models (LLMs) are helping with programming. For example, it is possible to extract a wide variety of named entities (people, places and dates) from documents and sometimes to identify or locate them. Text analysis methods and processes are becoming

¹ A. GRAVES, J. SCHMIDHUBER. 2008. *Offline handwriting recognition with multidimensional recurrent neural networks*, in the proceedings of the 22nd International Conference on Neural Information Processing Systems (NIPS'08), 545552.

² See eScriptorium, accessed 27 February 2025.

easier to implement. Moreover, documents are no longer accessible only to specialist researchers, but much more widely to very different audiences, as well as to private companies seeking training data.³ This is especially true when we add the increased capacity for text translation.

Similar examples could be given concerning image or video analysis. It should be noted that, from a historian's point of view, this is no longer just a matter of faster access to a larger mass of documentation, but rather an increase in the amount of information extracted from automatically generated data. It may then be asked whether this is not a form of big data in history. However, before making observations based on this data, it is important to maintain a cautious approach to its construction and to take an interest in what these algorithms produce. The cases highlighted here may seem transparent, but this is obviously not so. Today, there is much talk of the production of fakes, fake events and fake quotes. Historians know that these are not new phenomena, and once again it is necessary to understand their social and technological conditions of production. Hallucinations result from the way AI models are trained on very large amounts of data, which may include biased or incorrect information and may sometimes come from virtual documents.

AI and history: between old questions and renewed scientific approaches?

Persistent questions about the computerisation of the discipline

Insofar as it can be applied to various documents (images, maps, texts and audiovisuals, etc.) covering all historical periods and geographical areas, AI provides massive processing capacity that paves the way for comparative and longitudinal studies. It allows structures and patterns invisible to the human eye to emerge from vast historical data sets. But these are only probabilities. Thanks to increasingly effective machine-learning algorithms, AI can identify subtle correlations and emerging trends that traditional methods of analysis would not be able to detect.

Looking at the new scales of documentation made possible by neural algorithms over the past decade, it would seem that the terms of a recurring debate within the historical community once again apply to the situation we have just described. With the rise in the digitisation of the printed collections of major heritage institutions as of the late 1990s, on the one hand, and the ever-increasing computerisation of our contemporary societies, on the other, has called for much analysis of the consequences of the massification of data for the work of historians. The 2010s were marked by predominantly transatlantic calls to embrace big data in history,⁴ and other “macroscopic” approaches.⁵



The re-emergence of algorithms based on neural systems over the past fifteen years has enabled a remarkable level of recognition to be achieved, one that has never been seen before.



³ For example: HTR-United, accessed 27 February 2025.

⁴ P. MANNIN, *Big Data In History*, Palgrave Macmillan, 2013.

⁵ S. GRAHAM et al., *Exploring Big Historical Data: The Historian's Macroscopic*, Imperial College Press, 2016

“
With the rise in the digitisation of the printed collections of major heritage institutions since the late 1990s on the one hand, and the ever-increasing computerisation of our contemporary societies on the other, there has been much analysis of the consequences of the massification of data on the work of historians.
 ”

As a result, the decade was characterised by a series of epistemological debates on the scale of corporuses and the nature of historical phenomena to be studied. One of the most emblematic cases has undoubtedly been the reception of Jo Guldi and David Armitage's *History Manifesto*, published in 2014, to which the French journal *Annales* devoted an issue entitled “Debating the Long Term” (*La longue durée en débats*) in 2015. In the field of historical studies, the expansion of digitised corporuses to include manuscript documents seems to make the article published in 2008 by journalist Chris Anderson, on the predicted obsolescence of the scientific method in the age of data overload, more relevant than ever. Recent developments in AI in the field of history are thus part of a critical approach that avoids the overly techno-centric methodology which emerged in the early days of the relationship between history and computer sciences in the 1970s and 1980s.

The mirage of history that is not only computer-assisted but actually reduced to statements and observations corresponding to the highest probabilities calculated by LLMs undoubtedly represents a risk. Yet, we must remain level-headed and not lose sight of the fact that deep learning algorithms remain, first and foremost, tools adapted to the instrumentation of structured and detailed analyses of past phenomena. From this point of view, we can only reiterate that it is possible to view AI as a means of improving the accuracy and efficiency of research by enabling historians to explore new perspectives based on renewed digitised corporuses. It is, however, imperative to remain vigilant about potential biases and to maintain human control over the quality of the information produced. Let us bear in mind that, given the current state of AI research, these are only probabilities. They can be a valuable aid to research for many situations of uncertainty inherent in historical subjects, but they still need interpretation.

Using AI to address targeted issues

Whether we consider the use of computer vision methods to detect specific patterns in images, or those derived from NLP, a number of collective research projects demonstrate such possibilities. This is the case, for example, with the *HORAE* project, conducted at the Institut de recherche et d'histoire des textes, which studies religious practices in the late Middle Ages through Books of hours (Christian prayer books). On a completely different tack, the *EyCon* project should also be mentioned. It aims to study colonial, imperial and international armed conflicts from 1890 to 1918, by aggregating photographs scattered across multiple archival collections and automatically enriching the data from these visual sources to expand research possibilities.

It should be noted that these expansions are also made possible by the openness and sharing of methods and models produced in the course of various research projects. Take, for example, the article entitled *The Augmented Social Scientist*, in which the authors use an LLM to annotate millions of texts.⁶ In addition to the results themselves, the researchers

6 Do, S., OLLION, É., & SHEN, R. (2022). The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy. *Sociological Methods & Research*, 53(3), 1167-1200.

have also published an eponymous Python package to enable the scientific community to reuse similar methods. In the same spirit of openness in AI-assisted scientific practices, the “AI Toolbox” and “Deep Learning Models and Tools” projects should be mentioned, which aim to share with the community several tools and models developed as part of work on digitised archives. The shared resources include, for example, freely reusable and modifiable scripts for exploring the vocabulary of a corpus of texts built from Gallica, as well as automatic image classifications based on the contents of large data repositories such as the Joconde database of French museum collections.

However, all these projects require both human and material resources. These can be more or less significant for the underlying models to be deployed and used. In the case of the shared resources mentioned above, it should be pointed out that once the volumes of data processed reach a certain size, it is necessary to use platforms with sufficient computing capacity. This often forces researchers to use solutions such as Google Colab, which inevitably raises the question of the financial and environmental costs of such approaches, as well as the autonomy of research vis-à-vis large private economic players.

Understanding the limitations and potential biases of algorithms

Today, approaches and research are obscured by monstrous investments in LLMs, which too often reflect a muscle-flexing contest between states. Viewing these technologies in this way turns them into anxiety-inducing geostrategic tools. We hope, nevertheless, to have shown that a hammer may also be used to hammer in nails, although we have only touched on a few aspects of a vast issue that affects all aspects of history.⁷

Historians need to arm their critical reporting with new skills tailored to putting the results produced by AI into perspective, as well as understanding the limitations and potential biases of the algorithms used. The limited observations here can only lead us to emphasise once again the persistency of questions concerning the relationship between historical research and AI, since the development of computing capabilities during the second half of the 20th century. Could this ultimately be a sign of the routine interactions between history and computer science? ■

⁷ On these points, see the article by PHILIPPE C. BESSE on Mediapart, <https://blogs.mediapart.fr/fylb/blog/140225/boycott-citoyen-face-aux-investissements-delirants-pour-l-ia-usage-general>, published on 14 February 2025, accessed on 27 February 2025.

The Ambivalence of Professional Journalism Faced with AI's “Surveillance Capitalism” and Influence

In the era of “surveillance capitalism” and influence, professional journalism finds itself caught between dependence on platforms and the dilution of its democratic role. This article explores these tensions in the face of artificial intelligence (AI).



Jérôme Valluy

is a senior lecturer in information and communication sciences at the Sorbonne School of Political Science.

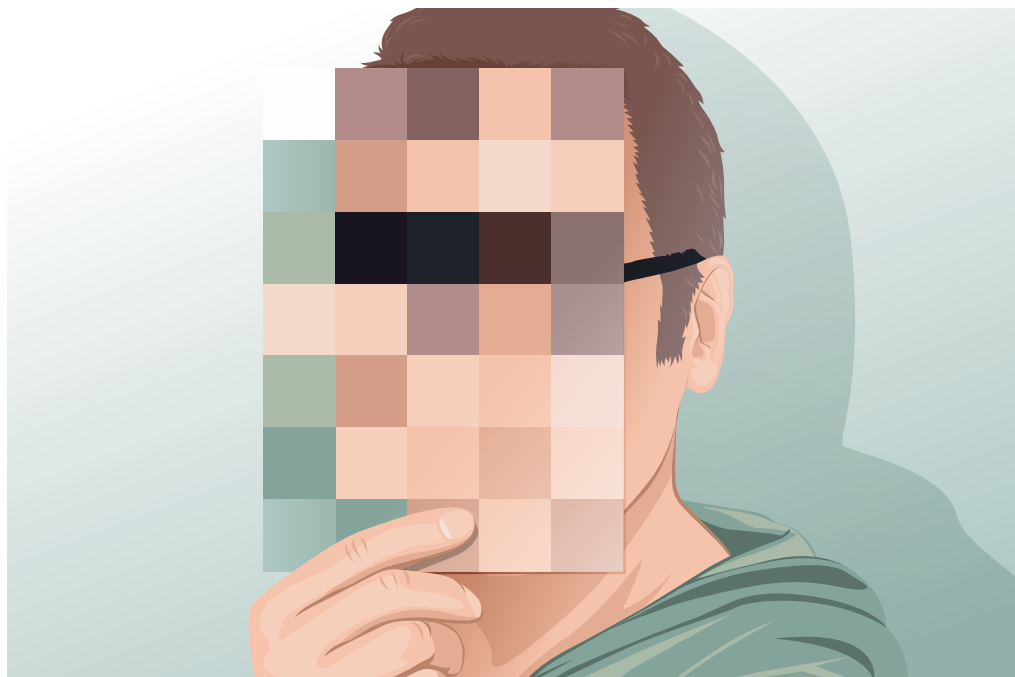
In *The Age of Surveillance Capitalism* (Zuboff 2018), AI is the backbone of the monetisation of personal data in the personalised advertising market. After an initial euphoric “digital revolution” (1995-2001), this system developed quietly in the United States between 2001 and 2011 and was only revealed later, with the Snowden revelations of 2013 and the Cambridge Analytica scandal of 2018 playing key roles. From a political science perspective, AI and surveillance capitalism have been intertwined since 2001, and even more so since the Cambridge Analytica trial revealed the extent of the manipulation of digital flows surrounding internet users. In looking at surveillance capitalism and influence, we can ask ourselves whether opening up AI to the public is a step forward or a step backwards for professional journalism – a pillar of democracy.

Professional journalism as an actor and victim of the digital revolution

It is not easy to talk about professional journalism in general: professional journalism is a mosaic of practices (local/national, investigative, data journalism, etc.) that warrant rigorous and thorough empirical studies. But in the face of changes as global and profound as those brought about by the digital revolution, these existing methods must be questioned, in as far as they blind us to what is happening. This mirrors Charles Wright Mills's insight (after World War II), when (in 1959) he criticised the “petty investigations” that neglect the general transformations facing political scientists and sociologists.

From this overall perspective, the role of professional journalism in the face of AI and surveillance capitalism is ambivalent: on the one hand, journalists, particularly in the United States, adhere to a very broad political consensus (from Democrats to Republicans) that masks the formation of surveillance capitalism. On the other hand, the professional media are also the main channels for revelations from whistleblowers, such as Edward Snowden and Frances Haugen, via *The Guardian*, *The New York Times*, *The Washington Post*, *Le Monde*, *Der Spiegel*, etc.

The portmanteau term “collution” (i.e. socio-economic *collusion* and editorial *dilution*), coined by Franck Rebillard and Nikos Smyrniaios, also illustrates the ambivalence of professional journalism, which is both a player and a victim of the digital shift. This change has taken place gradually,



without any overall strategy, as a result of adaptations by press companies. Digital technology has invaded the public sphere, giving increasing prominence to non-professional actors who challenge journalistic authority under the guise of freedom of expression and citizen counter-power. The result is a dilution of journalistic production and quality in a mass of non-professional content.

Historically, independent journalism has been built in opposition to established powers, as for example during the French Revolution. Alternative forms of digital journalism were initially perceived as an extension of this democratic tradition, giving them a certain legitimacy. Rather than fighting them legally (for unfair competition or copyright infringement), the profession has often come to terms with them. And such independent platforms have diverted advertising revenue away from the media, constituting an essential aspect of the collision between journalism and digital technology. But collusion and collusion go hand in hand: press shareholders negotiate the sharing of advertising revenue with digital giants rather than defending journalistic independence; more than 90% of journalists themselves share articles on social media, scrutinise public reactions, favour digital resources over field investigations and adapt to the logic of hashtags. How can we explain this collaboration, which paradoxically seems to undermine journalism itself? The explanation leads us to analyse a structural dependency.

Digital degradation of the journalism profession

In the 20th century, sharing a newspaper was common practice. Circulation was measured by the “circulation rate”, the ratio of readers to copies sold, generally limited to a few people per newspaper. With the digital revolution, this rate has exploded. No one anticipated that search engines and networks such as Google and Facebook would capture audiences and ad-

“
Digital technology has invaded the public sphere, giving increasing prominence to non-professional actors who challenge journalistic authority under the guise of freedom of expression and citizen counter-power.
”

“
The synthetic web (Ertzcheid) is self-perpetuating. Rather than curbing this trend, some media groups are exploiting it.”

vertising revenue without producing any original information themselves. Between 2005 and 2010, the widespread use of smartphones accelerated this change. In 2017, 96% of journalists used social media to promote their own content. From then on, the concept of a “circulation rate” disappeared. In 2021, *Le Monde* stopped measuring its readership in terms of the number of newspapers read, but in terms of subscribers on social media (more than 25 million in total).

Since 2001, surveillance capitalism has disrupted the media economy. Information is now “free” for users, financed by the monetisation of personal data and targeted advertising via AI, which is invisible and socially unquestionable. These advertising flows influence individuals without their knowledge. In less than fifteen years, this transformation has drained advertising revenue from traditional media to the benefit of American Big Tech digital platforms, whose enormous market capitalisation allows them to monopolise the global digital infrastructure.

Since 1995, professional journalism has been challenged, justifying citizen alternatives in the face of media concentration. The South Korean newspaper *OhmyNews* (2000) was born with the slogan “Every citizen is a reporter”. In 2004, Dan Gillmor published *We the Media: Grassroots Journalism by the People, for the People*, and while his *Bayosphere* project failed in 2005, it paved the way for other initiatives. In France, *AgoraVox* was launched in the same year, motivated by growing mistrust of professional media. The term “citizen journalism” became popular in the late 2000s, driven by the increase in Internet connections and the rise of amateur contributions. Olivier Tredan (2007) referred to it as an “ambiguous practice that is spreading like wildfire”, while Patrice Flichy (2010) analysed this as the *Consecration of the Amateur (Le Sacre de l’amateur)*. For their part, the social sciences have offered contrasting interpretations. Patrick Champagne (2006) viewed such citizen journalism as a democratic advance, making possible the emergence of the “collective intellectual”, as imagined by Bourdieu. Conversely, Nicolas Pelissier and Serge Chaudy (2009) denounce citizen journalism as “technopopulism”. Michel Mathien highlighted the return of “amateurs” and propagandists, historically opposed by professional journalism.

Then, the surveillance of journalists has become inexpensive with spyware such as *Pegasus* and *Predator*: as of 2011, *Forbidden Stories* lists 180 journalists who have been targeted. In 2024, Amnesty International denounced the inaction of states when the trade in spyware was used against journalists in Armenia, India, Serbia and the Dominican Republic. *The Washington Post* has, moreover, pointed out that *Pegasus* allows real-time geolocation, as illustrated by the assassination of Mexican journalist Cecilio Piñeda, 15 days after he was entered into the software’s database. More generally, *Reporters Without Borders* has counted 1,787 journalists killed since 2000.

The proliferation of artificial AI-generated publications since 2023 has undermined journalistic credibility. While the media uses AI to automate certain productions, the release of ChatGPT (2022) has accelerated this phenomenon. The *synthetic web* (Ertzcheid) is thus now self-perpetuating. Rather than curbing this trend, some media groups are exploiting it. In 2024, *Le Monde* and Springer signed an agreement with OpenAI to provide their data in exchange for revenue. Conversely, *The New York Times* (NYT) filed a lawsuit against OpenAI in September 2023 for the unauthorised use of its articles. An agreement would have been possible, but OpenAI, having

already obtained data via Springer and *Le Monde*, no longer saw any benefit from taking information from the NYT. These strategic choices are dividing the profession.

Journalistic dependence on digital sources and resources

Despite collapsing revenues and deteriorating working conditions, collaborations between the media and digital giants continue. This dependence affects both media bosses and salaried journalists, although their short and long-term interests vary.

Faced with the rise of Big Tech, the heads of the major media outlets, rather than defending copyright and fair competition, are negotiating a share of advertising revenue via “neighbouring rights”. This stems as much from an immediate financial imperative as from an imbalance of power vis-à-vis Google, whose audience is crucial for the press... and from a lack of support by governments. Debates on copyright crystallised after the rise in the 2000s of news aggregators (*Google News*, *Yahoo News*, *Bing News*) and social networks (*Facebook*, 2004; *Microsoft* invested in 2007). From the 2010s onwards, publishers attempted to obtain remuneration for their content, leading to lawsuits, and in 2012, Google was accused of copyright infringement by French publishers, though no judgement was handed down. Instead, an agreement was reached via the Press Digital Innovation Fund (FINP), with Google paying €60 million to avoid regulation. Germany introduced a neighbouring right (2013), but Google circumvented the law by refusing to negotiate with Springer, which eventually gave in and granted licences. In Spain, a similar law led Google to shut down *Google News*. These tensions culminated in 2014, prompting the European Union to adopt the neighbouring rights directive in 2019. However, five years later, the situation remains deadlocked. Claudia Cohen (*Le Figaro*) has noted that despite agreements with Google and Facebook, the French press is facing refusals from Microsoft and X (formerly Twitter) to establish information-sharing contracts.

The dependence of salaried journalists on the Internet can be explained by the evolution of their professional practices. Since 1995, digital tools have transformed their profession, a change that became even more pronounced after 2001. Platforms, blogs and networks have become essential for monitoring, editorial collaboration and investigation. Individual blogs offer journalists freedom of expression outside the traditional editorial framework. The internet has replaced more costly methods (polls, vox pops, long-term investigations) as a source of information, but has increased financial and functional dependence. Platforms such as Wikileaks (2006) have expanded this resource. In 2024, hashtags are still perceived as indicators of opinion, despite their manipulability via AI. Finally, social media allows journalists to develop a personalised audience, sometimes larger than that of their newspaper, which reinforces dependence and subjects them to digital influences.

These transformations deprive readers of the essential functions of newsrooms: prioritising information, following up on topics over the long term, and ensuring the reliability of information, etc. The expansion of AI in society appears thus to be a step backwards, and the defence of professional journalism, more than ever, a fight for democracy. ●



Debates on copyright crystallised after the rise in the 2000s of news aggregators (*Google News*, *Yahoo News*, *Bing News*) and social networks (*Facebook*, 2004; *Microsoft* invested, 2007).



All bibliographical references and sources

cited in this article can be found in Chapter 5 of Jérôme Valluy, “Humanité et numérique(s) – De l’histoire de l’informatique en expansion sociétale... au capitalisme de surveillance et d’influence (1890-2024)” [*Humanity and Digital Technology – From the History of Computing in Societal Expansion... to Surveillance and Influence Capitalism (1890-2024)*], TERRA-HN-éditions, Collection HNP, 2024: 2nd edition (expanded and updated): 15 Sept. 2024, 421 p.: <http://www.reseau-terra.eu/article1347.html>



The Uses of Large Language Models in the Social Sciences

Like machine-learning methods, large language models (LLMs) are transforming empirical research in the humanities and social sciences (HSS), by extracting structured data from raw texts, analysing sentiments, and conducting documentary research. Thanks to their natural language processing capabilities, these models offer new perspectives for analysing vast textual corpuses that were previously difficult to exploit. But such uses also have significant limitations; integrating LLMs into the methodological tools of the humanities and social sciences cannot be done without a critical understanding of how they work.



Machine-learning methods are transforming empirical research in the social sciences by offering new tools, particularly for predicting and exploiting data sources that were previously difficult to use, such as language. In recent years, the rise of large language models (LLMs) has represented a major advance in natural language modelling, both in terms of understanding and generation. In the context of research, these models offer the opportunity to automate certain tasks while reducing costs, including for text-based prediction, document similarity analysis, and data collection. However, many questions remain about their use, particularly due to the presence of bias, the difficulty of accurately assessing their level of uncertainty, and their lack of interpretability. In this context, this article provides a brief introduction to LLMs, explaining how they work, presenting some applications in the social sciences, and highlighting certain limitations in their use, in order to provide food for thought on the conditions for their application.

Clément Gorin

is a professor
and

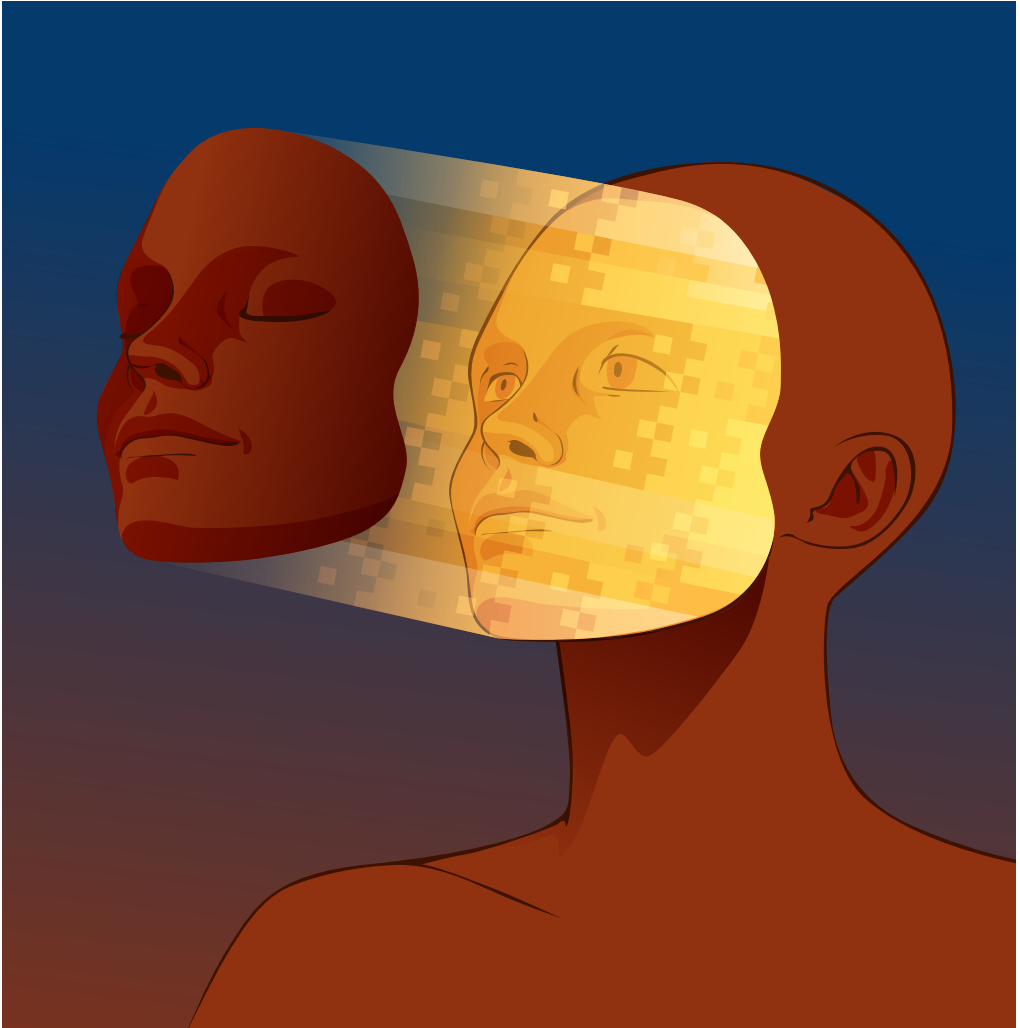
Thomas Renault

a senior lecturer, both in
economics at the
Centre d'économie
de la Sorbonne (CES)

What is an LLM?

LLMs are a family of machine-learning models designed to process natural language. These versatile models are based on neural network architectures called “transformers” and are characterised by a considerable number of parameters, estimated from large text corpuses using a method known as self-supervised training. To understand how LLMs work, it is essential to understand the structure of natural language. Natural language can be represented as a sequence of words,¹ and has two fundamental dimensions: semantics, which assigns meaning to the message and enables its interpretation, and syntax, which organises words according to grammatical rules and ensures the structural coherence of sentences. This duality makes automated language processing particularly complex, notably due to the absence of a numerical representation for the semantics of words and the need to capture the numerous syntactic interactions – sometimes between distant words in a text – that contribute to their meaning.

¹ Depending on the application, these units can represent words, sub-words or individual characters.



To address these challenges, neural networks calculate numerical representations of language in the form of contextualised vectors, known as “embeddings”. These vectors project words into a latent numerical space where proximity reflects semantic and syntactic similarities (Bengio et al., 2003; Le and Mikolov, 2014). For example, the numerical space will assign similar values to words used in comparable contexts, thus reflecting their semantic proximity, while distancing words that appear in different contexts. Each dimension of the vector encodes a specific aspect of meaning, which may correspond to an abstract concept or a characteristic shared between several words, although these dimensions are not directly observable. In terms of syntax, these representations also incorporate interactions between words, reflecting both their order of appearance and their hierarchical relationships. These relationships can be simple, such as grammatical rules, or more complex and abstract, such as analogies, as well as temporal and causal structures that contribute to the coherence of the text. Language models can learn these



To understand how LLMs work, it is essential to understand the structure of natural language.



representations by predicting a masked word from those surrounding it (Devlin et al., 2019). This task requires the model to develop a deep understanding of the semantic and syntactic dimensions of language.²

Among language models, the success of the transformer architecture (Vaswani et al., 2017) is based on a flexible and dynamic mechanism called “attention”,³ which allows these contextualised vectors to be calculated efficiently. Functioning as a question-and-answer system, this mechanism allows each word to interact with those around it in order to identify relevant associations. Thus, if the question posed by a word is answered by the preceding words, part of the meaning of the latter is integrated into the representation of the target word. In a transformer module, several “attention mechanisms” coexist, allowing words to ask various questions and obtain as many answers. Finally, the architecture consists of a series of these modules organised hierarchically, allowing language to be represented at various levels of abstraction. The first modules capture basic interactions such as frequent co-occurrences and basic syntactic structures, while the advanced modules represent more global and abstract concepts such as theme, emotion, or narrative structure.

What are their applications?

LLMs are considered to be foundation models, i.e. pre-trained architectures that have a general understanding of language and can be adapted to various tasks, sometimes without additional training. This section focuses on generic LLMs, without necessarily including those with conversational modules for chatbot applications.⁴

One initial application involves analysing the sentiment expressed in financial tweets in order to assess whether investor opinion is positive, negative or neutral towards a stock, market or economic trend (Renault, 2017). This process is usually based on manual annotation, which is often a time-consuming and costly task, sometimes requiring the expertise of specialised annotators. The use of LLMs reduces this dependency by relying on a pre-trained model that only needs to be adapted to a specific task. This process, called transfer learning, involves replacing the model’s output module with another one specific to the data distribution, like classification into categories such as positive, negative or neutral.

“In terms of syntax, these representations also incorporate interactions between words, reflecting both their order of appearance and their hierarchical relationships.”

² This is representation-based learning, where the model is trained on a secondary task aimed at acquiring high-quality representations. Furthermore, this approach is known as semi-supervised, as it uses raw text to automatically generate input and output data. In the case of generative models, prediction is made by determining the next word based on the preceding words (Radford et al., 2018). During inference, this same mechanism allows the model to generate a response in an autoregressive manner, using the question as the initial context.

³ LLMs rely on a specific attention mechanism called “self-attention”. Furthermore, this mechanism is formulated to exploit parallel computing, which allows the model to be trained on large text corporuses.

⁴ Conversational LLMs offer many other practical applications for accelerating certain daily research activities, such as interactive discussion to generate feedback, article summarisation, text correction and translation, or assistance with computer code writing and mathematical derivations, particularly with a new generation of so-called “reasoning models”. However, it is essential that researchers have the necessary knowledge to validate the quality of the results obtained. For an in-depth presentation of these applications, readers can consult Korinek (2023).

The model parameters are then refined on a task-specific sample, resulting in a high-performance model with a minimum of annotated data. Another application is to measure the similarity between documents using *embedding* vectors and distance metrics. Neural representations allow consistent and structured distances to be defined by capturing the semantic and syntactic relationships between texts. Thus, two documents can be identified as similar, even if they contain different words, a different sentence order or varying lengths. For example, Kelly et al. (2021) apply this method to the analysis of technology patents to identify disruptive innovations – patents that stand out from previous work while strongly influencing future developments. Textual distances make it possible to measure the novelty of a patent by comparing it to its predecessors, and to gauge its influence by assessing its similarity to patents filed later. By combining these two dimensions, this approach quantifies the impact of innovations and makes it possible to track the evolution of technological waves over the long term.

A final application concerns data collection. A specific architecture of generative LLMs, known as Retrieval-Augmented Generation (RAG), makes it possible to efficiently exploit vast document databases in order to extract relevant information. Unlike traditional generative models, whose knowledge is limited to the data acquired during training, RAG combines text generation with information retrieval from an external document database. This approach combines the flexibility of language models with greater accuracy in responses, as it relies on external, verifiable sources rather than simple probabilistic generation. RAGs thus simplify the use of specialised databases, such as historical archives or scientific publications, while significantly reducing the risk of errors.



A specific architecture of generative LLMs, known as Retrieval-Augmented Generation (RAG), makes it possible to efficiently exploit vast document databases in order to extract relevant information.



Under what conditions should they be used?

The use of pre-trained LLMs implies a loss of control over the data used for their training. These statistical models tend to replicate or even amplify the biases present in their training databases, which can lead to biased or discriminatory representations, particularly when these data lack diversity (Manvi et al., 2024). Furthermore, if the database used in the application is freely accessible, it is possible that the model has already been trained on the research sample. This can lead to overfitting, where the model memorises the data rather than extracting general trends from it, thus distorting inferences and compromising the validity of the results. To limit these risks, it is recommended to use open-source LLMs whose training data is documented and whose updates are clearly dated.

Another problem lies in the difficulty of accurately quantifying the uncertainty of LLM predictions. Unlike traditional statistical models, they do not provide confidence intervals for their predictions.⁵ This lack of uncertainty can lead them to produce erroneous predictions with excessive confidence. For example, LLMs are trained to reproduce the dis-

⁵ Several techniques inspired by Bayesian approaches can be used to estimate confidence intervals for parameters and predictions. They are based either on repeated sampling or on explicit modelling of uncertainty, taking into account both the variance of the parameters and that of the data. However, this second approach requires doubling the number of parameters in the model.

“
Unlike humans,
these models do
not understand
language
semantically, but
rely on statistical
correlations
derived from
training data.”

tributional structure of language, which can lead them to generate false but plausible representations, rather than rigorously accurate ones. One way to manage this uncertainty is to compare predictions to an external validation sample – one that was not used during training – and to explicitly model the structure of prediction errors (Ludwig et al., 2025). Finally, another challenge in certain applications is the lack of interpretability of LLMs. This opacity results from the complexity of their mechanisms, which rely on a considerable number of parameters interacting in a non-linear manner. This makes it difficult to trace precisely how a model constructs its representations and generates its predictions. Unlike humans, these models do not understand language semantically, but rely on statistical correlations derived from training data. As a result, their representations of language do not correspond to ours, which complicates their interpretation. Much work is being done to interpret the internal representations of models or to align them with those of humans, but this work mainly applies to architectures that are simpler than current LLMs.

Powerful tools for a wide range of tasks

LLMs open up new perspectives for social science research by facilitating language analysis, information extraction and prediction from textual data. Their flexibility and adaptability make them powerful tools for a wide range of natural language modelling tasks, while limiting the need for manual annotation. However, their use raises major methodological challenges, particularly in terms of bias, quantification of uncertainty and interpretability. Thus, for rigorous application in research, their use must be based on the fundamental principles of empirical validation and transparency of training data. ●

RÉFÉRENCES

- BENGIO, Y., DUCHARME, R., VINCENT, P., & JAUVIN, C. (2003). A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, 3, 1137-1155.
- LE, Q., & MIKOLOV, T. (2014). Distributed Representations of Sentences and Documents. *Proceedings of the 31st International Conference on Machine Learning*, 1188-1196.
- KORINEK, A. (2023). GENERATIVE AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, 61(4), 1281-1317.
- MANVI, R., KHANNA, S., BURKE, M., LOBEL, D., & ERMON, S. (2024). Large language models are geographically biased. *Proceedings of the 41st International Conference on Machine Learning*, pp. 1-16.
- LUDWIG, J., MULLAINATHAN, S., & RAMBACHAN, A. (2025). Large language models: An applied econometric framework. *National Bureau of Economic Research*, N° w33344.
- KELLY, B., PAPANIKOLAOU, D., SERU, A., & TADDY, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3), 303-320.
- RENAULT, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, 84, 25-40.
- VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., & POLOSUKHIN, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

-00:76:12:41:25



-00:36:12:41:08

-00:54:12:41:17

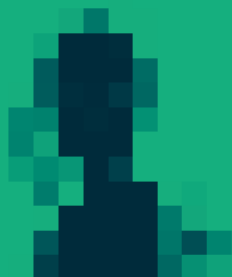


-00:22:12:41:34

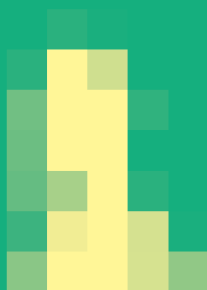
-00:47:12:41:00



-00:09:12:41:46



-00:06:33:00:45



-00:00:57:41:22



-00:00:42:41:55



00:00:00:00:00



00:00:42:41:55



00:00:57:41:22



00:06:33:00:45



00:09:12:41:46



00:36:12:41:08