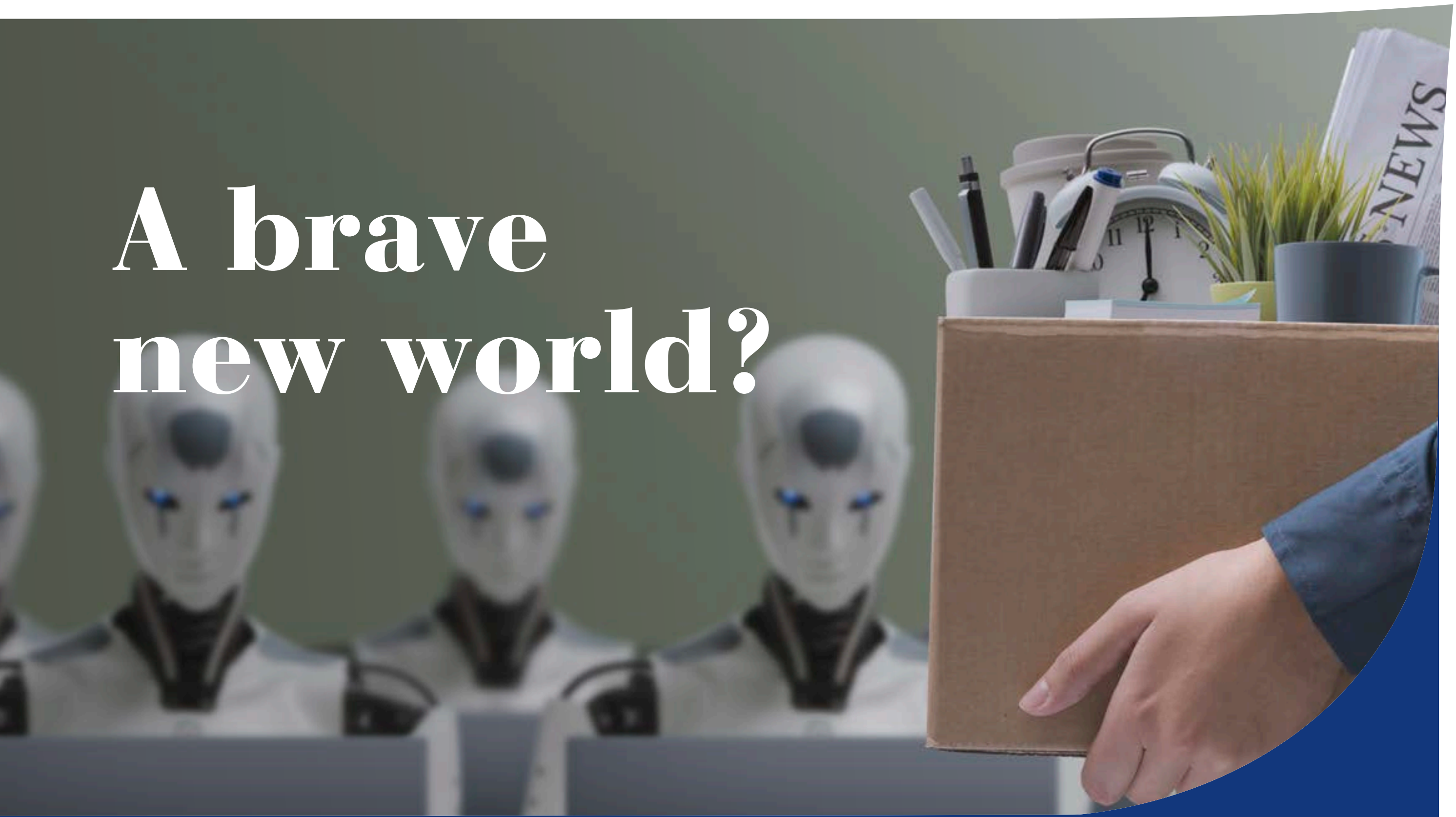
A brave new world?



Recent AI breakthroughs have raised fears about job losses in the knowledge economy, and revived concerns about sophisticated profiling that can intensify polarisation. Nick Pelosi and Ross Teverson explain how we are engaging with companies to help mitigate these risks.

Setting the scene

Since the release of ChatGPT in November 2022, artificial intelligence (AI) has captured the popular imagination. It has also become a daily talking point in the press, with excitement and anxiety evident in equal measure. This is because, while AI has the potential to dramatically increase the productivity of businesses and transform many aspects of our daily lives, it also brings with it considerable risks.

As generative AI tools proliferate, the biases they sometimes reflect may perpetuate stereotypes and stall progress towards greater equality.[1] There is also the possibility that jobs may be lost to AI at a faster rate than new employment opportunities can be created, particularly in certain sectors. Adding to these concerns is the challenge that, while most people agree that unbridled AI deployment could lead to significant unintended societal harms, there is currently little agreement on how to regulate it.[2]

While AI algorithms are now seamlessly embedded into our daily lives, we are only just beginning to grasp the implications for companies and civil society. Generative AI has made a significant leap forward with the launch of ChatGPT, but these large language models may randomly generate false or misleading information, known as AI hallucinations.[3] The problem is that the casual reader has no way of identifying what is true and what is false.

> **In the absence of effective regulation, EOS has been engaging on the business and wider societal impacts of AI since 2017.**

For further information please contact:

Ross Teverson
Sector: Technology
ross.teverson@FederatedHermes.com

Nick Pelosi
Theme lead: Digital Rights
nick.pelosi@FederatedHermes.com

[1] Humans are biased. Generative AI is even worse – Bloomberg

[2] The EU is leading the way on AI laws. The US is still playing catch-up. Everyone accepts that AI is dangerous. Agreeing on what to do about it is a different story – The Guardian

Given that social media can be used by hostile actors to spread problematic content, undermine democracies and influence the outcomes of key elections and referenda, what are the implications of wholesale adoption of generative AI? The question is a pressing one given that policymakers remain substantially behind the curve on regulating the use of AI. This risks creating a free for all in which unverified content is served up daily across social media platforms.

In the absence of effective regulation, EOS has been engaging on the business and wider societal impacts of AI since 2017. In 2022, we consolidated our approach to engagement on this topic under the wider sub-theme of digital rights, which we define as human rights specific to digital products and services.

Our Digital Rights Principles[4] set out our core expectations of companies on AI. These explain that companies should ensure robust governance and policies for AI. Companies should disclose the range of purposes for which they use algorithmic systems; explain how they work, including what they optimise for and what variables they consider; and enable users to decide whether to allow them to shape their experiences. Companies should take action to eliminate unintended racial, gender, and other biases in algorithms, including those recommended by the EqualAI Checklist to Identify Bias in AI.[5]

Our Investor Expectations on Responsible AI and Data Governance white paper, published in 2019,[6] sets out a full engagement framework based on six principles as follows:

Trust

Companies should earn trust by educating users on their rights to data privacy, and give users control and the right to consent to the use of their data by providing fully free choices.

Transparency

Companies should be transparent about tracking methods in the full value chain and disclose how they measure the robustness of data governance and the fair and safe use of AI. Companies should inform users when their data is being used for scoring and screening purposes.

Action

Companies should thoroughly explore and make all reasonable efforts, in good faith, to avoid unintended consequences such as data and process bias, which may lead to discrimination.

Integrity

Companies should demonstrate integrity in the treatment of customers, suppliers, and users. They should avoid user manipulation, including approaches that encourage addiction, such as shopping, gaming and device addiction that goes beyond the limits of targeted advertising. Companies should have risk disclaimers about addiction and consider providing users with an opt out from targeted advertising.

Accountability

Companies should establish a clear accountability system internally and externally within their AI development and application ecosystems. There should be an appropriate due diligence process for supply chains and third-party access. Companies should build systems that allow for auditability and put in place appropriate insurance where possible.

Safety

Human safety is of paramount importance, especially when it comes to accessing critical services, such as water, electricity and healthcare, or control of transportation such as autonomous vehicles. Companies should demonstrate that their AI applications put human safety as a priority over profit and revenue.

This report also sets out a 'three lines of defence' model for trusted AI implementation. Each category of the assessment is mapped to the principles and analytical framework (legal and financial factor analysis and salient social impact analysis) that we highlighted in the April 2019 paper. Ethics sits at the core and is the first line of defence. Risk, governance and audit form the second line of defence. Responsible use of AI embedded in strategy and operations is the third line of defence.

Updates to our approach

Given the rapid pace of new developments in AI, we have continued to review and evaluate our engagement approach. We will continue to engage on AI as a human rights issue and we are closely exploring the overlay with two other themes: human capital management, and wider societal impacts.

Human capital management

In 2017, McKinsey estimated that these technologies could displace 15% of the global workforce by 2030 – the midpoint of its scenario range.[7] These technologies impact employment, wages, and working conditions through the displacement effect, in which they replace workers or suppress wages, and the productivity effect, in which they enhance workers' efficiency or create new jobs.[8] The loss of millions of jobs could exacerbate social inequalities and increase civil strife. However, optimists believe that improved productivity will offset the displaced jobs, as was the case during the industrial revolution. We expect companies to demonstrate how they will manage the transition, including planning for any job losses, such as through re-training.

3 https://www.techopedia.com/definition/ai-hallucination#:~:text=An%20AI%20hallucination%20is%20where,outputs%20from%20large%20language%20models.
4 https://www.hermes-investment.com/uploads/2022/04/5a8aadeb037fb131b1889c3f6b1a85aa/eos-corporate-digital-rights-principles-04-2022.pdf
5 https://www.equalai.org/assets/docs/EqualAI_Checklist_for_Identifying_Bias_in_AI.pdf
6 Investors' expectations on responsible artificial intelligence and data governance | Federated Hermes Limited (hermes-investment.com)
7 mgi jobs lost-jobs gained_report_december 2017.pdf (mckinsey.com)
8 A new study measures the actual impact of robots on jobs. It's significant. | MIT Sloan

We also expect companies to disclose if and how their use of AI, automation, and robotics are impacting their workforce. Disclosure should provide quantitative and qualitative information about the jobs displaced and other impacts to employment, wages, and working conditions. It should describe the company's policies and practices for managing impacts, such as ensuring that workers are given sufficient notice and/or priority for other open positions. And the company should demonstrate evidence of retraining, upskilling, and other forms of financial or technical support for workers impacted by the transition.

Wider societal impacts

An evolving issue, which requires more consideration, is company lobbying on AI regulation. Strong AI regulation that mitigates unintended risks inevitably slows innovation to some extent. However, major players in AI, including Google and OpenAI, recognise that consistent, global, and collaborative regulation may be necessary to avoid a race to the bottom and mitigate extreme risks.[9]

We may therefore spend more time seeking to understand the approach that companies are taking to AI regulation, with an expectation that they support a regulatory approach that helps to mitigate the risk of societal harms and any subsequent financial impacts on businesses.

Additionally, companies should go beyond simply complying with regulation when it comes to deploying responsible AI through human rights due diligence throughout the business, including for capital expenditure on AI or other emerging technologies. These procedures should show how risks to wider societal outcomes are considered in business decision making as well as the clear no-go areas or restrictions being imposed on technologies.

CASE STUDY

DBS Group

Our engagement on AI with DBS started in 2020 with a letter to the chair. We recently met data governance and compliance representatives of DBS face-to-face in Singapore for an AI-focused discussion. We sought to understand the bank's approach to AI governance, how it implements its ethical AI principles and how it manages potential societal impacts.

Responsible data use is governed at various levels across the company, such as business units, dedicated functional committees, management and the board. The company ensures effective implementation of its responsible data use framework, PURE (Purposeful, Unsurprising, Respectful, Explainable), through a combination of training, approval mechanisms, compliance and audit assessment, debates at committees and fit-for-purpose reviews.

It puts significant emphasis on building educational content and seeks the board's direction proactively. It plans to establish structured reporting on AI to the board. To ensure that innovation is balanced with stringent governance, it socialises ideas at the senior level, then tests models in a controlled environment before deployment.

We asked whether the company's advocacy is focused on achieving regulatory consistency across the markets in which it operates. While differences in policy across its markets pose challenges for ensuring consistency, it continues to take proactive steps to stay abreast of changing policy and shifts in societal norms across the jurisdictions in which it operates.

We discussed how it assesses materiality and the potential for significant social harm across its use cases. It evaluates ethical use and unintended bias against customers, employees and the disadvantaged, and scores use cases against a number of dimensions to determine the materiality of each. It is working on testing use cases with higher materiality and ensuring recourse and the explainability of its models as a next step. We welcomed the progress made to date and encouraged the company to continue to increase the transparency and oversight of this rapidly evolving field.

Sonya Likhtman
Theme co-lead: Natural Resource Stewardship

Judi Tseng
Sectors: Financial Services, Technology

[9] Perspectives on Issues in AI Governance – Google, Governance of superintelligence – OpenAI

CASE STUDY

Royal Bank of Canada

We have been engaging with the Royal Bank of Canada (RBC) on AI governance since 2020. RBC's market position means that it can play a pivotal role in establishing responsible policies for the use of AI.

In Q3 2020, we asked RBC to publish AI ethical use principles and encouraged it to assign board level responsibility to a director. In early 2021, we met with the company and were encouraged to hear that, while the board did not have a designated AI expert, the topic was discussed at board meetings.

RBC had also commissioned an internal white paper highlighting ethical AI principles and said that its practices aligned with those of the Borealis AI Institute, a research centre created by the bank focusing on responsible AI.

However, we emphasised the value of making these principles publicly available to alleviate any stakeholder concerns.

In Q1 2022, we continued to press the company for more clarity on board and senior management level oversight of AI. After receiving a shareholder proposal on AI governance in 2023, the company said that the board receives regular reports on initiatives involving the use of AI, and the risk committee reviews significant and emerging risks, including those related to the adoption and use of AI. This, along with the expertise of directors, led us to believe that the company had sufficiently addressed board level AI oversight.

We also view the company's work on the Respect AI pillars (which identify robustness, fairness, model governance, data privacy and explainability as critical for responsible and safe AI), along with disclosure outlining RBC's approach, as emerging best practice.

Emily DeMasi
Sector co-lead: Financial Services

Technology and financial services

EOS currently engages on over 60 objectives or issues that relate to AI. We recognise that AI advances human development, but there is also the potential for misuse. Perhaps unsurprisingly, over half of the companies engaged are in either the technology or financial services sectors.

The technology sector is leading the development of AI itself while also using AI for numerous purposes. Tech companies play a huge role in influencing users' behaviour or contributing to social segmentation, exerting significant control over the media consumed.[10] For example, social media companies use AI to curate, rank, and recommend online content, to deliver targeted advertising, generate search results, and feed users political news. This can lead to the development of so-called echo chambers, where a user is unwittingly digesting only one side of a story, reinforcing their existing views. This fuels the polarisation of political and cultural opinions. Increasingly, people are driven to take more extreme positions and the consensus is lost.

We ask companies in our engagement programme to build trust in responsible AI through various methods. For example, we expect companies to demonstrate that their business models do not incentivise problematic content, to include specific examples of AI deployment in their human rights impact assessment, and to provide disclosure of the policies and processes they use to enforce child age restrictions where relevant.

In the financial services sector, AI deployment is also widespread. Specific applications include risk management, chatbots, virtual assistants, underwriting, fraud detection and algorithmic trading. One of the key issues in AI deployment is the potential for racial and gender bias. We have engaged on this issue to gauge how companies are thinking about it. We have also asked companies to publish the ethical AI principles that their AI models follow and to consider conducting a bias assessment.

Outlook

As AI deployment becomes more widespread in sectors beyond technology and financial services, the scope of our engagements will extend to new areas. However, we anticipate that an emphasis on AI governance and ethical use principles will continue to form the foundation of our approach. Meanwhile, understanding and mitigating the impacts from a human capital management and wider societal impacts perspective will become increasingly important.

[10] Unintended racial, gender, and other biases have been identified within algorithms and can lead to inequitable outcomes.

Federated Hermes

EOS

Federated Hermes

Federated Hermes is a global leader in active, responsible investing.

Guided by our conviction that responsible investing is the best way to create long-term wealth, we provide specialised capabilities across equity, fixed income and private markets, multi-asset and liquidity management strategies, and world-leading stewardship.

Our goals are to help people invest and retire better, to help clients achieve better risk-adjusted returns and, where possible, to contribute to positive outcomes that benefit the wider world.

All activities previously carried out by Hermes Investment Management are now undertaken by Federated Hermes Limited (or one of its subsidiaries). We still offer the same distinct investment propositions and pioneering responsible investment and stewardship services for which we are renowned – in addition to important strategies from the entire group.

Our investment and stewardship capabilities:

- **Active equities:** global and regional
- **Fixed income:** across regions, sectors and the yield curve
- **Liquidity:** solutions driven by four decades of experience
- **Private markets:** real estate, infrastructure, private equity and debt
- **Stewardship:** corporate engagement, proxy voting, policy advocacy

For more information, visit **www.hermes-investment.com** or connect with us on social media:

EOS001195 0015864 09/23.