OECD PRINCIPLES ON ARTIFICIAL INTELLIGENCE: BUILDING ETHICAL AI

CAPCO

THE FUTURE. **NOW.**
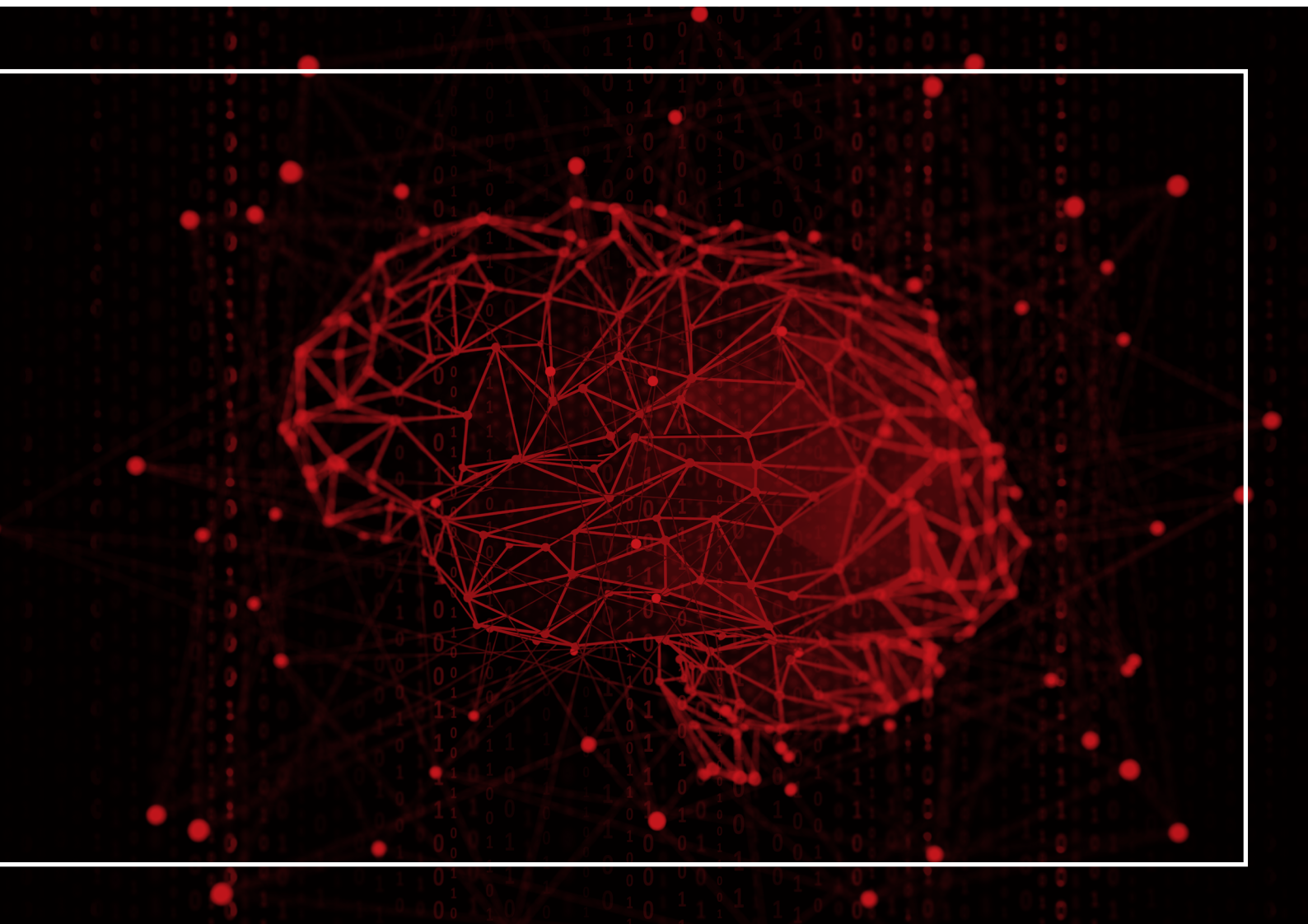
ABSTRACT

In May 2019, Canada and forty-one other countries adopted the artificial intelligence (AI) principles outlined by the OECD. Although not legally binding, these principles will likely form the basis for future international standards and national legislation. Given the rapidly evolving nature of AI, promoting these principles requires organizations to start early and to stay current with technological developments.

" Although not legally binding, OECD AI principles will likely form the basis for future international standards and national legislation."

- Kevin Marshall Ph.D, Senior Data Scientist, Capco

ARTIFICIAL INTELLIGENCE UNCHAINED

Artificial intelligence (AI) promises to make a positive impact on humanity by helping solve a range of problems from self-driving cars to early cancer detection but left unchecked AI has proven its ability to learn to reproduce the worst aspects of its human architects. Our biases. Our history of discrimination. In 2017, Amazon scrapped development of an AI recruitment tool it had hoped would help improve the efficiency of finding qualified candidates[1]. The AI had learned from its training data that a successful candidate was male (there remains a significant gender gap at top U.S. technology companies[1]), and so it learned to downgrade resumes which contained the word 'women.'

THE OECD PRINCIPLES ON ARTIFICIAL INTELLIGENCE

Recently, member countries of the OECD approved a set of guiding principles meant to serve as the first intergovernmental standard on AI[2]. The spirit of the principles calls on all AI actors to promote and implement: "inclusive growth, sustainable development, and well being; human-centered values and fairness; transparency and explainability; robustness, security, and safety; and accountability." As financial institutions continue to grow their AI footprint, they should proactively consider the impact these principles have on their enterprise AI strategy.

These principles are not legally binding; however, the OECD has a rich history of providing the foundational framework for national legislation. For example, the OECD Privacy Principles, adopted in 1980, formed the foundation for the Data Protection Directive[3] adopted by the European Union in 1995 and then superseded by the General Data Protection Regulation (GDPR) in 2018[4].

While the principles on AI serve as a flexible standard for a rapidly evolving field, national regulations are not far behind. Understanding what is on the horizon and building solutions in consideration of these policies will allow financial institutions to stay ahead of the curve.

Explicitly, the OECD Principles on AI often refer to methods "consistent with the state of art[5]." In leveraging this rapidly evolving field, AI actors are responsible for building trustworthy and transparent solutions that are agile enough to stay current. The question of transparency and fairness in many machine learning (ML) models remains an open and active area of research. Financial institutions should be leveraging the best methods available from the field to promote the principles on AI. However, compliance is not a once-and-done process; as the field matures and as humanity unifies on the expectations of AI, responsible actors must ensure adequate approaches are taken to ensure fairness and transparency.

IMPLEMENTING THE PRINCIPLES ON AI

There are two key aspects to consider when implementing the OECD AI principles in a financial institution. First, creating an environment supported by a robust governance model is critical; one of the five principles focuses on accountability— all AI actors must be accountable for the proper functioning of AI systems based on context and their roles. Next, although many technical challenges hinder the true realization of the AI principles, solutions are actively in development in both industry and academia, and financial institutions must stay aware and utilize methods consistent with the state of art. Below, we discuss current considerations and common approaches for two of the technically challenging principles:

Model fairness

Ensuring a fair (unbiased) AI system is not as simple as just removing protected attributes (e.g., gender) from the data used to train models. In the case of the Amazon recruitment tool, not only was the word 'women' used to downgrade resumes but implied factors including whether the candidate attended an all women's college or applied characteristically masculine verbs were found to be culprits. Implementing AI fairness generally consists of two steps: measuring fairness using pre-defined metrics, and then mitigating possible biases.

Just as fairness is a multifaceted social concept with no single simple definition, fairness metrics come in a wide variety. Generally, these metrics fall into two classes: group fairness, where you compare statistical measures across population groups defined by the protected attribute, and individual fairness, where you compare similar individuals between each other. Likewise, you can apply bias mitigation algorithms in many ways (e.g., on the data, the models, or the predictions). Metrics and algorithms should be selected appropriately depending on the use case. The concept of AI fairness has evolved beyond just academia, and open-source tools are available for practical assessments.

Transparency

As AI models grow in complexity, there is a characteristic trade-off between performance and explainability. The 'black box' nature of many of these models can foster a lack of trust and acceptance of the technology. Furthermore, under the principles, individuals adversely affected by an AI system should have the opportunity to challenge its outcome based on straightforward and easy-to-understand information about the features used as well as the underlying logic.

Financial institutions should ensure that not only are the appropriate methods of explanation being leveraged, but that customers are aware of when they are being affected by an AI system and that the proper process for transparency is in place. One way to improve the explainability of an AI system is to use a combination of simple and sophisticated models. While the sophisticated models may be able to process complex data and relationships, the simple models offer a rationale that is interpretable in the eyes of stakeholders.

Conclusion

Financial institutions possess a wealth of sensitive customer information that can be used to fuel AI which improves customer experience; however, it also means they bear a greater responsibility to ensure the ethical use of the information. Successful realization of the OECD principles on AI will take coordination and accountability at all levels.

At Capco we have been guiding our clients on how to leverage a practical approach to AI implementation. The topic of responsible use of AI has been top of mind to our financial services clients and being considered as operating models are rolled out. Banks have proceeded cautiously with heightened management attention and risk mitigation by their various risk and compliance control parties as AI models and techniques are developed and implemented.

We'd love to hear your thoughts and challenges in responsibly realizing the benefits of AI, for more information, please contact **joanna.lewis@capco.com**.

REFERENCES

1. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

2. https://www.oecd.org/going-digital/ai/principles/

3. http://oecdprivacy.org/

4. https://link.springer.com/article/10.1007/s00439-018-1919-7

5. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

AUTHORS

Kevin Marshall Ph.D, Senior Data Scientist
Joanna Lewis, Partner

CONTRIBUTOR

Hoan Dang Ph.D, Data Scientist

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward.

Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and asset management and insurance. We also have an energy consulting practice in the US. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

To learn more, visit our web site at **www.capco.com**, or follow us on **Twitter**, **Facebook**, **YouTube**, **LinkedIn** and **Instagram**.

WORLDWIDE OFFICES

APAC	EUROPE	NORTH AMERICA
Bangalore	Bratislava	Charlotte
Bangkok	Brussels	Chicago
Hong Kong	Dusseldorf	Dallas
Kuala Lumpur	Edinburgh	Houston
Pune	Frankfurt	New York
Singapore	Geneva	Orlando
	London	Toronto
	Paris	Tysons Corner
	Vienna	Washington, DC
	Warsaw	
	Zurich	**SOUTH AMERICA**
		São Paulo

WWW.CAPCO.COM

CAPCO
THE FUTURE. NOW.