Smith institute

INNOVATING IN AN UNKNOWN REGULATORY ENVIRONMENT:

Developing the artificial intelligence of the future

2023 // DR GEORGINA LANG

Smith institute

Dr Georgina Lang
Senior mathematical consultant

INNOVATING IN AN UNKNOWN REGULATORY ENVIRONMENT: DEVELOPING THE ARTIFICIAL INTELLIGENCE OF THE FUTURE

The next few years will be an exciting time for artificial intelligence, and we can expect to see AI becoming even more commonplace.

Generalist knowledge-based AIs like ChatGPT have captured everyone's attention, but progress will be rapid in specialist AI too. AI-backed algorithms for forecasting renewable output could help the transition to a greener electricity grid, and AI-powered processing of medical images could see doctors making better and quicker diagnoses.

Alongside the greater prevalence, we will be entering an era of greater regulation to protect the safety and rights of citizens. The regulatory outlook is evolving quickly but is still uncertain. This makes it a challenging time for AI developers as the AI products of the future will need to comply with as-yet-unknown regulations.

The evolving regulatory landscape

AI regulations would define mandatory requirements applicable to the design and development of certain AI systems before they are placed on the market, to ensure that people's safety and rights are not at risk. For example, an AI-assisted surgery robot must not be detrimental to a patient's health, and an AI algorithm to filter CVs must not reproduce or exacerbate past biases.

Currently there is little AI regulation, but change is coming, with different nations at different stages on the regulatory journey. The EU is proposing the first-ever legal framework on AI, taking a risk-based approach with stronger requirements for high-risk applications, and in March 2023 the UK government published a white paper, setting out their pro-innovation plans.

Similarly, other countries including the US, Canada and Australia are taking the first steps towards their own regulatory frameworks. Indeed, New York city has passed a law requiring companies using AI software for hiring decisions to have the technology audited anually for bias, effective from July 2023. Of course, any AI used in multiple geographic locations will have to adhere to all relevant regulations, meaning that the strictest regulations will need to be complied with. However, we do expect that all regulatory frameworks will be governed by similar principles.

Regulation is ultimately a positive step as it paves the way for AI to be better incorporated into everyday life, governed by considerations of ethics and safety. However, it does also pose challenges: how can you develop a compliant AI product when you don't know what regulations you need to comply with?

Building future-proof AI?

As the regulatory landscape evolves, no developer can know for certain what the future of AI regulations will be.

> ***We can apply both common sense and guidance from the principles that have been published by governments to ensure that products are developed in the spirit of the regulations to come.***

The EU's proposed 'risk levels' provide helpful information to assess the risk of your AI product, and hence guide the level of regulation required. A sensible first step is to consider which level your product will fall into. At one extreme, the EU is to ban applications of AI that exhibit a clear threat to the safety, livelihoods and rights of people, for example social scoring by governments. At the other extreme, minimal and low risk applications – such as chatbots and spam filters – will be subject to only transparency requirements or no regulation at all.

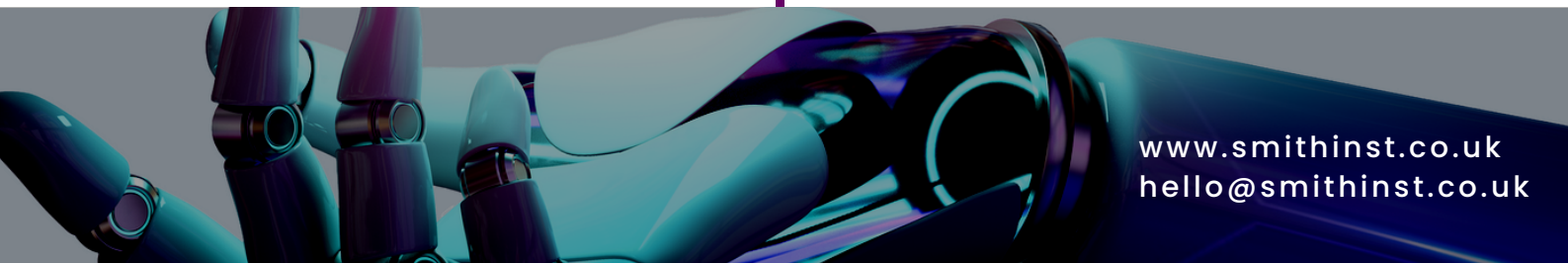
The middle ground of high-risk applications, such as an AI for management of critical infrastructure or credit scoring, is where regulations will need to be carefully considered.

An approach to future-proofing your AI during this time of uncertain regulation is to ensure your product adheres to the guiding principles of a relevant government's proposed policy. For example, the UK government's guiding principles provide a useful checklist of development good practice that should be followed by those creating the software:

- **Safety, security and robustness**. This principle spans considerations of risk-based testing, data adequacy and software security. Any AI should be subject to regular testing to ensure it remains fit for purpose and relevant to the current situation. This may include testing with curated data sets to probe challenging scenarios and edge cases, assessing the sensitivity of outputs to inputs, and performance testing to ensure that the overall results are adequate. The level of assurance should be proportionate to the risk of the application.

- **Appropriate transparency and explainability.** Many AI algorithms are black box, meaning that users cannot understand their inner workings. One approach to mitigate this is to build an explainable layer using a technique such as SHAP, giving users some insight into the factors driving the output. An alternative is to limit your algorithm to an interpretable model, whose decision-making process can be followed exactly.

- **Fairness**. Organisations should be aware of the risks of bias in training data, and consider whether training data – be it real or synthetic – is representative of the inputs the AI will see in practice. Encompassed within fairness are legal considerations such as GDPR for personal data and copyright laws.



- **Accountability and governance.** Organisations should ensure that there is oversight of AI systems, with clear lines of accountability. As data scientists, we can contribute by ensuring that the AI methods, assumptions and risks are well communicated to those who are responsible for governance.

- **Contestability and redress.** For it to be possible to contest the outcome of an AI, robustness and explainability are key. An important consideration is whether AI should make automated decisions or provide guidance to human decision makers. While automated decision making is suited to many contexts such as low-risk and high-throughput scenarios, some high-risk applications may be more suited to using AI outputs as guides for human operators.

These helpful principles encapsulate much of the good practice which responsible AI developers are already striving for.

We don't know the specifics of future regulation. However, we can use nascent frameworks of different jurisdictions to anticipate their coming approaches. **The typical outlook looks to be flexible and pro-innovation, where AI leaders, developers and regulators can collaborate in parallel**. Abiding by moral and practical principles, bearing in mind the risk of each specific AI application, will set up AI developers for positive and constructive engagements with regulatory bodies.

This will sow the seeds for future uses of AI that are both exciting and safe.

If you would like to understand more about implementing responsible, trustworthy AI, get in touch:

hello@smithinst.co.uk
www.smithinst.co.uk

Smith institute

THE QUEEN'S AWARDS FOR ENTERPRISE: INTERNATIONAL TRADE 2021