

## Next-Generation AI and Advanced Autonomy

Despite impressive advancements in AI, we are still a long way from systems that can perform complex activities autonomously in real-world domains. Outside tightly controlled environments, we are not yet able to build systems that exhibit the flexibility, adaptability, and awareness required to perform those tasks which we would prefer AI systems to take on rather than humans. Also, the currently dominant paradigm of training systems from scratch on each new problem – with data and compute requirements increasing dramatically as we attempt to tackle ever more complex problems – is becoming unsustainable. Without combining and reusing components flexibly “on the fly”, we will likely soon hit a hard barrier both in terms of the capabilities of AI systems and their sustainability.

To address this challenge, our vision for **advanced autonomy** is based on effecting a step change in terms of how we develop, integrate, adapt, and interact with AI. We expect that this will require a concerted effort to develop next-generation AI methods by working across all areas of AI, where key new advances will likely come from a range of methods and subfields that many researchers are already working on, including (but not limited to):

### *New ways of dealing with data*

Moving beyond the currently practice of training systems on large amounts of data will likely require advancing techniques such as one-/few-shot learning, the combination of different data sources and use of new forms of data, or new ways of using high-level knowledge in combination with observational data.

### *New ways of assembling systems*

To overcome the focus of current systems on very narrow, often single-shot tasks, and enable on-the-fly combination of individual components in novel modular and expandable systems, methods such as meta-learning, multiagent coordination, or neuro-symbolic architectures will likely be needed.

### *New ways of adapting systems*

Enabling systems to adapt to different contexts of use will require exploring new forms of adaptation, for example through human instruction and demonstration, but also to enable adaptation to the needs of diverse user populations, using novel lifelong learning, task generalisation/abstraction, and crowdsourcing techniques.

### *New ways of interacting with systems*

Increasingly complex AI systems will likely be programmed and configured by developers and users at runtime in novel, more natural, and incremental ways. Adjustable autonomy, mixed-initiative techniques, human-AI interaction/communication and novel AI programming techniques will likely play an important role here.

### *New ways of understanding and using autonomy*

Whatever levels of autonomy we strive for, future AI will have to complement human activity meaningfully, exploiting the strengths of artificial and human intelligence. Ensuring we pursue responsible and beneficial innovations will require extensive exploration of the ethical, socio-technical, and human aspects of autonomy.

## Laying Foundations

We have recently embarked on a Turing 2.0 project called “Enabling Advanced Autonomy through Human-Collaboration”, a collaboration between the University of Edinburgh, the University of the West of England, and the University of Oxford that aims to lay the foundations for a future UK-wide major initiative on this theme. This project involves over 15 robotics and vision, natural language processing, knowledge representation, multiagent systems, cognitive science, and human factors researchers working together to demonstrate the potential of novel, integrative AI methods to advance autonomy. Focusing specifically on new forms of human-AI collaboration through naturalistic interaction as a new design methodology, we will work on enabling embodied AI agents to learn new skills from human instructors at run-time through a continuous cycle of instruction, explanation, and feedback. Alongside this core technical research, we will also explore the human and ethical dimensions of establishing advanced levels of autonomy through new forms of human-AI collaboration.

## A Call to Action

While the project will pursue one specific approach to exemplify the potential of integrative, collaborative efforts to break new ground in terms of unlocking a step change in autonomy, our wider ambition is to build a UK-wide community of researchers working across all areas of AI and interested in developing a long-term research agenda for the wider area. This will involve mapping existing research capability in relevant areas, bringing researchers together to shape our shared vision, and developing a research roadmap – all with a view to establishing a major national initiative that will allow us to pursue this vision over the next ten years.

If you are interested in participating in our activities, please complete the form at

<https://forms.office.com/r/ZkQhvTNYzY> and/or contact [Michael Rovatsos](#) (project lead) directly.