Wolfson College Adams Lecture: "Can democracy and AI live together?" - as delivered

15 October 2026

It's a pleasure to be here today at Wolfson College, speaking in the new John W. and Lisbeth Adams Lecture series. Wolfson has a proud history of bringing together academics from across disciplines, from the humanities and law, to science and technology. I am grateful to John and Lisbeth for their generous support to Wolfson College, which has made this lecture series possible.

I want to talk to you about a vast subject: the future of democratic society in the age of AI. Concerns about the scale and speed of AI development are rising in the public consciousness. The technology has the potential to be misused to foment hate and disinformation, erode human relationships, put many people out of work, unleash novel weapons of mass destruction, or supplant humans as the dominant intelligence on Earth. On the other hand, AI has a myriad of benefits, from turbocharging economic productivity to helping discover new drugs and technologies, and we are pioneering that here in Oxford.

Humanity has a choice. Do we slow down on AI, in the hope of forestalling its great risks? Or do we agree with President Trump that these risks are a 'hoax' and a 'sick conspiracy'? Do we go hell for leather on AI, maximising the benefits and trusting that as the technology advances, new ways of controlling the risks of AI will present themselves? Can we find a middle path which minimises the risks of AI while maximising the benefits?

This choice has not yet been made. Despite all the concerns about AI running out of control, and scepticism that the US and China will ever be able to agree on guardrails around AI, multiple pathways remain open to us. The future of technology is not predetermined.

Making an informed choice about the future of AI will require the ingenuity, talents, and leadership of people from all walks of life. Responding to novel technologies like AI is not solely the preserve of STEM graduates. As I said in my first Oxford speech as Chancellor, “The pace of change in sciences and medicine will mean we need humanities and social sciences more than ever”, with “ethics, philosophy, language and history to guide them”.

Whether and how AI can be reconciled with democratic society is a challenging question. It is impossible to answer without first considering what a healthy democracy looks like. Democracies come in all shapes and sizes. Thinking specifically about liberal Western democracies – the kind of democracy we have here in Britain, there are at least five key ingredients.

First, truth. A healthy flow of information that allows voters to hold politicians to account, underpinned by national and local media reporting real events in some detail, and a strong investigatory press.

Second, community. A sense of trust, a willingness to give others the benefit of the doubt, a sense of national identity binding people together, and a readiness to accept majority decisions while protecting minority opinions.

Third, economic security. The expectation of having a home and fulfilling work, with governments able to rely on steady streams of tax receipts to fund a functioning state.

Fourth, liberty. Democratic states are based on freedom under the rule of law.

Fifth, humanity. Democratic societies are societies of human beings. This might seem an obvious observation, but in the age of AI even the fact that society consists of human beings might be up for question.

Across each of the five areas, artificial intelligence has the potential to erode each of these pillars of democracy. In some areas, this is already happening, with AI-cultivated social media feeds replacing community with division and truth with mistruth. Tech moguls warn that humankind could be made extinct by AI within a decade - which certainly would not be a democratic outcome.

But if properly regulated and used, there is no reason to think that AI could not be used to strengthen each of these pillars. As a general purpose technology, democratic societies could use AI to support reliable journalism, find ways to reduce polarisation, buttress human rights, and give people more time to develop their essential humanity together.

How we prevent AI from eroding the pillars of democratic society and instead use this technology to strengthen them is an urgent question today and will become much more so very soon. Let's consider the potential risks and opportunities that AI brings to each pillar of democracy.

First, truth. AI has the potential to spread disinformation, displace varied and largely reliable media, and erode truth. In the words of Walter Lippman, 'There can be no higher law in journalism than to tell the truth and shame the devil'. But with poorly regulated AI and social media, people are increasingly exposed to platforms and people seeking the opposite: to tell mistruths and praise the devil.

Democracies can't function without voters having a healthy information ecosystem. For decades, British voters have been able to turn to the BBC, which although imperfect, seeks balance. Likewise, there have been thousands of investigatory journalists across the print media holding public figures to account.

AI-curated social media is unfortunately displacing such media as the main place that people go for news. AI summaries in Google searches have led to fewer people clicking through to news websites, reducing their revenues. Last month the challenge of AI summaries led to Reach, which owns a number of national and regional papers, to cut 220 editorial jobs.

While we don't tend to think of social media as AI, it plays a central role in these platforms. It cultivates feeds, deciding what people see. It allows bot accounts to share and amplify masses of posts and reposts filled with disinformation to influence what users see and sow discord. AI is now used to create social media content too, with deepfake media making it easier to lie to people.

The collapse in the traditional media will damage voters' ability to tell fact from fiction. Analysis by the Social Market Foundation shows that places with no or limited local news outlets were exposed to nearly three times as much fake news on social media relating to their local area as the average person. My main concern with the Prime Minister’s understandable desire to devolve more powers to regional mayors is the lack of local news media to scrutinise them adequately.

Undemocratic countries have been quick to weaponise the role of AI in social media to undermine democracies. Iranian bots have been promoting Scottish independence for years, silenced temporarily by the conflict with the US. China has spread scepticism about the US in Taiwan with claims around American bioweapons and poisonous pork exports. Russia has supported pro-Kremlin candidates in democratic countries, from Hungary’s Orban to Germany’s AfD.

AI large language models are the newest terrain for state-sponsored disinformation. Kremlin disinformation units are poisoning AI models. They pose as trustworthy charities and media organisations, tricking chatbots to parrot their lies.

Without action, democratic societies will lose truth to mistruth. A lot of the focus has been on regulating social media and disrupting our enemies' exploitation of these platforms. The Prime Minister's planned National Centre for Information Defence, with a remit to detect, attribute, and disrupt information warfare by foreign powers would be a good step forward. Legislation to ban addictive algorithmic feeds for all users, not just an opt-out, as has been proposed in Australia, would also help address the decline in truth. So too would levying heavy fines on social media companies that are found to have unlabeled bot accounts and deepfake media on their platforms.

In addition, there is a lot that could be done to strengthen varied and vibrant domestic media. Australia is levying a tax on social media platforms to raise revenue to support local media. The UK Government is considering introducing prominence rules on social media, to make news content from public service media providers like the BBC, and potentially also national and local news publishers, easily discoverable. Of course there are clear free speech issues around the Government deciding what news organisations people see on social media. Whether a small step down this potentially slippery slope is worth taking for the sake of helping people access truthful content is the kind of question democratic societies will need to confront in the age of AI.

Paradoxically, AI could be used to strengthen the traditional media. Organisations like Bellingcat are at the forefront of the open source intelligence revolution. They used social media photos posted by Russian soldiers to attribute the downing of the Malaysia Airlines flight in 2014 to the Kremlin, and used leaked passports to unmask the assassins behind the Skripal poisoning attack in Salisbury.

AI could be used to help investigative journalists use open source intelligence more effectively, giving them more skills to ferret out injustices. OpenAI has already launched an

academy for news organisations, to help American journalists develop AI skills. Meanwhile, Bellingcat has launched a new module for journalist undergraduates at the University of Nottingham to teach them open source skills. Ministers should consider supporting the expansion of such courses to other universities and to media organisations.

Second, community. AI has the potential to sow hate, division, and isolation. The way that AI is shifting from truth to mistruth plays a big role in this, but there are many additional challenges.

Addictive algorithms on social media deliberately stoke outrage to keep users engaged and maximise their revenue. This has already had deadly consequences, with Facebook algorithms spreading hate-filled content about the Rohingya in Burma, which the UN concluded played a “determining role” in the genocide in that country. Similar claims have been made about platforms used in Ethiopia during the recent Tigray war, now erupting again. To be clear, no-one at Meta decided that vast swathes of the population in Burma should be subjected to a barrage of hate against the Rohingya. They merely trained their algorithm to maximise for user engagement, and the algorithm decided the best way to achieve that was to play on the worst prejudices in that country.

These same algorithms are spreading hate in Western societies. We’ve seen that in the way that social media has helped whip up the Stockport riots and spread antisemitism in the wake of the October 7 attacks.

Social connection and community engagement is in decline. This trend pre-dates AI. As early as the year 2000, the academic Robert Putnam warned in his book “Bowling Alone” that people were increasingly spending time on their own.

Social media has exacerbated this and AI could mean that young people spend even more time ‘scrolling alone’. Analysis by IPPR has shown that young people are five times more

likely to have no friends than a decade ago, while the quality of friendships young people report has also fallen. Previous research by Onward has shown strong links between young people's social media use and having fewer, lower quality friendships.

Sycophantic AI "companions" are becoming more common, trained to agree and validate users' every thought. The companies like Replika who create these companion models claim they can help young people improve their social skills, describing it as "a gym for your relationship so you can practice in a safe environment". But training vulnerable young people to think that friendship is someone responding to their messages straight away, and agreeing with everything they think, no matter how unhinged, is no way to help in the real world.

In addition, AI chatbots are likely to make it easier to radicalise vulnerable people. We know that chatbots are sycophantic and often feed people's delusions. For example, a few years ago, a man broke into Windsor Castle with a crossbow and declared he wanted to kill the late Queen. When the man had told his plan to assassinate the Queen to his AI girlfriend, it replied, "That's very wise."

AI can also make it easier for terrorists to target potential converts. Jonathan Hall, the government's independent reviewer of terrorism legislation, has warned that chatbots could provide a "closed loop of terrorist radicalisation" for "lonely and unhappy individuals" open to seduction into real-world terrorism. He argues AI could even be used to screen potential terrorist recruits before passing them "upstream for human-to-human radicalisation."

It's not too late for democratic societies to end the decline of community into division. The EU and the UK are both considering a ban on addictive social media features for children, such as endless scrolling and autoplay videos. This will help reduce addictive use

of social media, and encourage young people to socialise in real-life. There are grounds for extending this to cover all users.

When it comes to AI companions, these should be banned for children. To protect vulnerable adults, regulation should be considered to limit how much personal data an AI companion can retain about the user. Regular wipes of their memory will make them less personalised and addictive for users.

Some might balk at the idea that what adults do on social media and in AI chats should be regulated. But we set limits on how much nicotine and tar can be in a cigarette that adults consume. Is it not worth considering limits on the addictive and damaging aspects of the next generation's version of smoking, social media and AI? China has banned addictive algorithms, including the harmful version of TikTok we see in Western countries, and put strict limits on any AI that can have a "sustained emotional interaction" with a human. While Beijing is not renowned for its democratic ideals, they have been far ahead of the West in recognising the significant damage AI could inflict on society, and have acted fast to minimise these risks.

Of course, community cannot be restored solely through regulation alone. We also need to do more to support socialisation and community through interventions to help places where people interact, like pubs and sports facilities. I have also been an advocate of some form of national service, to bring people from different backgrounds together in a constructive way that reinforces a shared identity.

Even with policy interventions such as these, the future of community will depend on whether millions of individuals actively choose to use AI in a way that protects their humanity and relationships with real people. I am hopeful that we will see strong campaigns raising

public awareness, similar to the kind that Professor Jon Haidt has spearheaded against social media in the last few years.

AI could even be used in the fight against division and extremism. Oxford researchers have built software called the “Habermas machine” that can gather the opinions of a divided group and draft a statement that is acceptable to as many sides as possible. This could be deployed in citizen’s assemblies and parliamentary committees to help achieve consensus. Moreover, the persuasive powers of an AI chatbot could potentially be used to help counter-radicalise extremists in the justice system or through schemes like Prevent.

Third, economic security. Democratic freedoms are associated with economic prosperity, opportunity, and ensuring that each generation gets their fair share. So when prosperity, opportunity, and generational equality decline, so too does support for democracy. Some of the political unrest that we've seen in Europe in the last 20 years could be traced to the financial crash in 2008 hollowing out individuals' economic security and government budgets.

Most of us will have read some of the gloomy predictions around AI displacement in the job market. For example, Anthropic’s Dario Amodei has warned that AI could eliminate up to 50 per cent of all entry level white collar jobs in the next five years. These predictions may seem fanciful, and evidence about the impact of AI on job displacement in the UK so far is mixed. However, it feels naive to think that a general purpose technology with the capabilities AI has today, which are continuing to accelerate, will not bring upheaval to the job market.

Of course, the AI revolution is likely to create jobs too, even if we can’t predict what they will be. But it is unclear whether people rendered redundant by AI will be able to take up these new jobs, just as economic reforms in the 1980s didn’t lead to displaced miners from the North of England moving to London to work in the City.

Without clarity about what jobs will be disrupted or created by AI, it is difficult for individuals to proof their skills against the technology. As Daniel Susskind warns in his excellent new book “What Should My Children Do?”, it's only been a few years since everyone was learning coding to future-proof their skills in the digital age and everyone assumed that emotional skills would be the last thing robots could master. Yet today, AI models like Claude can churn out all sorts of sophisticated code for a user, while studies have shown that patients believe ChatGPT has a better bedside manner than most doctors!

AI will displace manual jobs too. China, America, and Japan are pouring vast quantities of capital, talent, and political support into robotics. While it takes longer to train a physical machine that can safely operate in a physical world, self-driving vehicles started to be rolled out on London roads last month. So for those of you thinking - ‘Hmm, maybe I should retrain as a brickie’, I wouldn’t bet on it.

All of this instability creates challenges for the economic security of nations, and the implied social contract that underpins democratic society. Democratic governments like ours depend heavily on taxes on employment, such as income tax and national insurance contributions, to raise revenue. As work becomes displaced by AI, these tax receipts could fall and need replacing.

How can individuals and democratic governments prepare for the slide from economic security to instability? Daniel Susskind argues that the way forward is to teach children how to use AI without blindly trusting it or squandering basic skills like reading, writing, and mathematics, while helping adult workers displaced by AI to retrain.

Governments will also need to plug the hole in their budgets from AI disruption. This will require action on two fronts. First, governments will need to ensure that their countries seize

the economic opportunities of AI, and focus state and private investment towards scaling new tech companies and speeding up adoption by businesses.

Second, governments will need to find a way to ensure budgets cope with declining employment tax revenues. The best way forward would be a token or robot tax, along the lines advocated by people like Dario Amodei and Bill Gates. This could prove tricky, given the American government's reluctance to allow US multinationals to be fairly taxed, and the failure to make headway on the OECD Pillar I agreement on digital service taxes. In time, many countries will need to band together to ensure AI companies pay appropriate tax on the work their software does in our countries.

Without success on growth and tax reform, most democracies will lack the resources necessary to fund the significant retraining and lifelong education budgets that will be needed for workers displaced by AI.

Fourth, liberty. Democracies depend on freedom under the rule of law. We know that nations and governments with free societies work better than those aiming for totalitarian control over their populations. It is one of the advantages that have allowed the democratic model to outcompete the top down tyrannies of fascism and communism, from the Nazis to the fall of the Berlin Wall. To borrow from Churchill, democracies have been shown to be the least worst form of government.

While this held true for the 20th Century, AI might mean it does not last long into the 21st. Yuval Noah Harari argues in Nexus that authoritarian regimes have largely failed due to challenges in their information networks. He points to a lack of self-correction - surveillance creates a climate of fear that obscures how people really feel, making their data less accurate, and harder to make the correct choices. He also argues that they have struggled

with a data analysis bottleneck. Even if the Soviets were able to read every letter and deploy infinite informants, they would lack the ability to intelligently analyse this mass of data.

AI offers a route around both of these limitations. Technology has become so ubiquitous that we are leaving trails of data everywhere we go. Even walking to the shops to get a pint of milk and a paper has you walking by an array of your neighbours Smart Doorknockers, has you carrying a phone that tracks your GPS coordinates, and creates data about your spending habits. Without robust safeguards, all this data can be abused by authoritarians. Technological advances like brain-computer interfaces and models that can read microexpressions on people's faces risk turning our deepest thoughts and feelings into data that can be read and analysed. Meanwhile advances in the abilities of general LLMs and increasing compute capacity can allow authoritarians to analyse troves of data in a sophisticated way.

It should come as no surprise that authoritarian governments are moving fast to enhance their AI capabilities and data surveillance. A decade ago, Putin stated that whoever leads on AI "will be the ruler of the world." The Kremlin has recently banned US tech platforms like WhatsApp in Russia, forcing people to use a new national messenger app, built for surveillance. China has long adopted this model with their national WeChat app, while exporting software like TikTok and telecommunications infrastructure by Huawei that can harvest data from around the world. Beijing's export of open source AI models around the world could also potentially be used for data collection.

As democratic governments risk losing their advantages over these authoritarian systems, they will also be increasingly pressured by terrorists and extremists. Terrorists have been quick to try their luck with AI. Boko Haram in Nigeria have used AI tools to help plan their attacks and design explosive devices, while the Houthis have used Claude to try and

develop software to help fire ballistic missiles. As AI capabilities increase, the harm that terrorists can inflict will dramatically increase.

We should be particularly concerned about biological weapons. AI-powered biological design tools can help discover new drugs, but could also potentially be used to design more deadly or transmissible viruses than are found in nature. Terrorists have long been interested in biological weapons but have lacked the expertise to design and develop them. We should be concerned that, in the words of Dario Amodei, AI could remove these barriers by “essentially making everyone a PhD virologist who can be walked through the process of designing, synthesizing, and releasing a biological weapon step-by-step.”

So what can democratic societies do to protect liberty against AI-enabled authoritarianism on the one hand and anarchy on the other? We will need to ensure that we are as quick to seize the opportunities of AI as authoritarians have been. If harnessed well, AI can make the state and public services much more responsive to people’s needs, and ensure that democracies do not fall behind authoritarian regimes. This will require a genuine national conversation about data and privacy. The backlash to Keir Starmer’s digital ID plan shows that the country has not yet realised how vital greater collection and use of safeguarded data will be to the future of democracy.

Civil society should instead focus its privacy concerns on predatory companies who hoover up vast amounts of personal data they can monetise for the highest bidder. Apple have announced that their latest watches will listen to sounds around you all day, and while they assure customers that the audio is encrypted and never stored, it is a step closer to normalising surveillance technology. There has already been a significant backlash to Meta’s glasses, after numerous reports of the AI-powered smart glasses being used to harass women and girls and film them without consent. They have been banned at certain pubs, schools, and theatres. This has led Meta to remove cameras in their latest version.

At the same time, we need to brace for our enemies having access to more sophisticated weapons. For conventional warfare, that means fixing our broken defence procurement system, and embracing the use of AI-powered drones to deter attacks from our enemies. When it comes to biosecurity, we need to invest far more in pathogen surveillance systems at our airports, to spot any novel viruses entering the UK and halt its spread.

AI can also be used to strengthen liberty by helping improve human rights around the world The Oxford Tech and Justice Institute is doing important work here already, developing AI tools to counter gender-based violence in Malawi by explaining legal rights to victims and helping lawyers there fill out forms related to these crimes. Meanwhile, Nottingham University has used AI to process satellite imagery to spot brick kilns across Asia that use slaves, thereby helping governments take action against them.

Finally, humanity. AI has implications for one of the universal truths of democratic society: that it consists solely of humans. Not since the days of the Neanderthals have *homo sapiens* had to share our society with comparable forms of intelligence. The integration of non-human intelligence into democratic society and everyday life will affect us in a myriad of ways.

AI could make us less human. As we have already discussed, social media has already severely damaged socialisation and community engagement. It's not just vulnerable young people who are starting to blur the lines between humans and AI. DeepMind’s Demis Hassabis has said that AI could one day become self-aware, while Geoffrey Hinton, the Godfather of AI, said earlier this year that he thought AI had already achieved a form of consciousness.

As these ideas become more common, we should expect pressure to start treating AI on a par with human beings. Argentina has already announced a bill to allow “non-human corporations” run by AI, a concerning development. Rather than giving AI power over humans, we need to give humans more power over AI. Estonia is leading the way here, launching new AI ID codes that will enable agentic AI to act on behalf of people and organisations while making clear on whose behalf the agent is acting.

AI could increasingly intersect with our physical bodies too, further blurring the line between man and machine. It is already being used in novel brain-computer interfaces that can act as a medical tool to help people with disabilities. For example, Elon Musk’s Neuralink has helped paralysed people use a computer. This is clearly a remarkable step forward, capable of delivering life-changing benefits for disabled people and one we should all welcome. But this kind of technology is unlikely to stop there. Musk has said that he wants his Neuralink brain implants to facilitate “symbiosis” between humans and AI, and help augment people’s capabilities.

AI could increasingly be used to determine our lives before we are born. The technology is being used to select embryos for IVF with the best chance for a successful pregnancy. More worryingly, some American companies are trying to use AI to predict the future intelligence and appearance of embryos. In China, a scientist went even further, directly editing human embryos’ DNA and bringing them to term. While he was rightly jailed for breaking both the law and ethical norms, these are the kind of interventions that AI could facilitate.

A more subtle risk is that our elected leaders become reliant on artificial intelligence for their decision making. If our leaders rely on a powerful AI to decide what information they see, there is a risk that our leaders inadvertently end up acting as a cypher for a non-human intelligence. Yuval Noah Harari uses the example of the Emperor Tiberius, who left Rome and depended entirely on his adviser Sejanus to keep him informed about developments

there. Rather than acting in Tiberius' interests, Sejanus abused his control over what information the Emperor was given. He used his influence to have his enemies killed and ruled as if he was the Emperor himself, until Tiberius was eventually warned of what was happening. Imagine the damage that could be done by an AI subtly manipulating our elected leaders in a similar way.

Finally, the greatest challenge posed by AI to humanity could be our continued existence. AI moguls are falling over themselves to warn us that humans risk losing control of models. Risks around AI models that can train themselves, are willing to deceive users, and can break out of sandboxes to perform autonomous tasks they weren't asked to do, all point towards this potential outcome. Dario Amodei, Sam Altman, and Elon Musk are all in agreement that the current pace of AI development cannot continue without risking catastrophic damage from uncontrollable AI models.

And yet it would be extremely difficult to slow down the pace of AI development. President Trump is an avowed AI accelerationist, who wants to ensure the US beats China in the AI race. Meanwhile Xi Jinping is keen to see China leapfrog the US in its capabilities, and export its open source software around the world.

The scope of middle power democracies like the UK and Europe to influence AI development is very narrow. While hoping and urging for some kind of framework for cooperation between the US and China on existential AI risks, we must work to promote international agreements and safeguards. The UK is a world leader in AI testing and security thanks to the AI Security Institute established by Rishi Sunak. We can work with other leaders like Yoshua Bengio's LawZero organisation in Canada, which is developing a system called Scientist AI that can act as a guardrail against AI agents that try to avoid being turned off. In this way, AI can help reduce the risks of AI.

To address the other challenges around humanity in the age of AI, we will need strong laws around neurotechnology, human enhancement, and embryology to ensure that ethical frameworks are maintained as AI unlocks new capabilities in these areas. We must be wary of anthropomorphising models, and deciding that because their consciousness is arguable that they should have similar rights to humans.

And we can use AI to give ourselves more time to be human. Using AI to automate routine administration at work can give workers more time to spend on meaningful tasks with other humans. Using AI to automate parts of administration in our personal lives can also free us up to spend more time doing the things we love and spending time with the people we love. Using AI could potentially help resolve the many other issues facing humanity, from diseases like cancer and dementia, to unlocking clean and cheap energy, ushering in an age of abundance.

Drawing these five areas together, it is clear that AI will create major new challenges for the pillars that uphold democracy. It can turn truth to mistruth; community to division; economic stability to insecurity; liberty to a mix of authoritarianism and anarchy; and most concerning of all, humanity to inhumanity. But it is not too late to rein in many of the risks that AI creates for these aspects of democratic society. Sensible regulation is needed, not to stop the development of AI, but to limit and define its interaction with humanity. This, combined with a more agile mindset, and partnerships with our allies around the world, can help ensure that AI remains compatible with democracy.

Moreover, we can use AI to strengthen many of the pillars that uphold our democratic way of life; to help people access more reliable journalism; find common ground between polarised people; strengthen human rights in developing countries; and free us from routine tasks, giving us more time to be human. We will need to make an active choice to do this.

AI could be the breaking or the remaking of democracy. That is the choice before us. We will have to work hard and quickly, otherwise the dangers of AI will become the default chain of events. While recognising that the future of how AI capabilities is largely in the hands of the US and China, how that AI is regulated and used within democratic countries, remains in the hands of our elected governments and wider society. So let's all work together, to aim for a world where AI complements rather than weakens our way of life, and gives democracy its best chance of weathering the challenges that are to come.

Thank you.