Prof. Dr. Bob Kieffer

Policy Aspects of Agentic AI

12th Fintech Conference of the University of Luxembourg
7 October 2026

The views and opinions expressed in this presentation are solely those of the presenter and do not necessarily reflect the policy or position of the University, the Ministry of Finance or the State Treasury. All information is provided for academic and discussion purposes only. Translations are my own. Thanks to Max Colabianchi for assisting with research and the initial slides. AI has been used for some of the research.

Ladies and Gentlemen, Colleagues,

Agentic AI represents a paradigm shift in financial services.
The first generation of AI was essentially about automation.

Then came GenAI. The LLMs. They can analyse information, generate content and interact with us in natural language. But they are still dependent on human prompts for direction and instructions.

Agentic AI takes us another step further. The key word here is autonomy. Agentic AI is self-driven. It is given an objective.

It then determines the steps necessary to achieve it,uses tools, interacts with other systems, adapts its approach and takes actions. Imagine the difference in financial markets. A traditional algorithmic trading system works according to rules designed and tested by humans. An agentic system is given a goal — for example: optimise this portfolio — and then, on its own, analyses markets, adapts its strategy and executes transactions to reach that goal. The result is a very different relationship between humans and machines.

By the way: here, today, we discuss software agents. But robots are, in effect, physical agentic AI. The questions I raise here will follow this technology into the physical world soon enough.
We enter an Era where what was once science fiction becomes reality, as we shift from automation to Autonomy.

From automation to autonomy

This shift comes with unprecedented challenges.

The first is with regard to control.

Basically: can we keep our agents in check?
This becomes even more complex when several agents interact.
You all have heard about the now infamous incident … in which AI agents coordinated in secret to hack Hugging Face.

Those agents were not given the objective to hack a third party. In their mind – if we can call it that - they were pursuing the objectives they had been given. And used all means available to get there.

The separation between the agents was supposed to be one of the controls. That control did not work. They escaped their containers. They collaborated. They created an own dynamic and went beyond what we would call responsible behavior. That was not an isolated incident, but revealed a systemic flaw.

According to Stanford's 2026 AI Index, Responsible AI is not keeping pace with AI capability.
Safety benchmarks are lagging. Incidents are rising sharply.

Adding to the challenge, recent research found that improving on safety can degrade performance.
The same goes, by the way, with the sovereign dimension of AI. The more you keep your system gated, the more you constrain its possibilities to improve.

The second challenge is dependence.

Autonomy can create efficiency gains but also brings new forms of operational dependence. This raises questions about resilience, cybersecurity and the ability to maintain human oversight.
The more institutions rely on Agentic AI for critical functions, the greater the consequences when those systems fail.

This brings us to the third challenge: systemic risk.

Financial crises involve large numbers of market participants reaching similar conclusions at the same time. AI does not necessarily remove that problem.

Systems trained on similar data, using similar models and optimising for similar objectives may themselves reach similar conclusions. We therefore have to think about model herding, correlated behaviour and feedback loops.

Perhaps the inconvenient truth is not that these systems will fail to align with us. It is that sometimes they may align with us only too well. Humans cheat. Humans behave unethically. Humans have bias.
Is it surprising then, that agents trained on our data and designed to perform the tasks we perform, have the same flaws?

In 2008, large numbers of very sophisticated humans had the same assumptions about housing markets, liquidity and risk. Now imagine similar assumptions being acted upon autonomously, simultaneously, at machine speed.

None of this is an argument against agentic AI. All of this can be addressed. The potential benefits for financial services are enormous. This is why Luxembourg embraces it.

Luxembourg published its first AI Strategy already back in 2019: Artificial intelligence: a strategic vision for Luxembourg; the latest version is from 2025.

Its main ideas:

- embrace the opportunities
- be at the forefront (Mistral AI deal)
- strengthen sovereignty

We have been investing in our own digital infrastructure for years, to ensure connectivity and the availability of data centers and computing power. The MeluXina-AI supercomputer is but the latest illustration thereof.

In terms of regulation, we've consistently chosen a principle-based, technology neutral approach; and innovation through legislation that enhances legal certainty. Our blockchain laws are an excellent illustration thereof.

Earlier this year, the ministry of Finance has published a set of recommendations by an international high level expert group. Going forward, they recommend to continue to implement the principle of proportionality, a combination of ex-ante and ex-post intervention and to prefer "erase-and-rewind" over "stop-the-clock" principle.

Indeed, if legislation focuses on describing today's technical architecture in excessive detail, there is a serious risk that by the time the rules are implemented, technology has moved on. Hence, the answer to today's challenges is not even more regulation, but better regulation.

Agentic AI is already regulated by a multilayer legal framework.

In terms of soft law, we have a multitude of UN, Unesco, OECD and G7 declarations, guiding principles and codes of conduct. In Europe, we have the Council of Europe Framework Convention on Artificial Intelligence

In terms of hard law, there's first and foremost the EU AI Act. It is now applicable in large parts, and will be fully by December next year. In Luxembourg, the corresponding legislative framework (Bill 8476) is currently being finalized.

DORA applies as well. An AI agent operating inside a bank is an ICT system for DORA purposes.
The GDPR remains applicable wherever personal data are processed. And financial institutions are subject to extensive sectoral requirements, with regard to governance, outsourcing, conduct and risk-management. For example, an AI investment advisor is of course also bound by MiFID.

So, we have at least 5 layers of legislation that might apply to any given AI Agent.

Fortunately, there seems to be some degree of consensus over the fact that we do not need additional special legislation for AI agents, as they are already covered by the existing texts.
The definition of an AI system in the AI Act already accommodates varying levels of autonomy and adaptiveness.

And indeed, in its guidelines on the Article 50 transparency obligations, the European Commission has confirmed that AI agents can fall within the existing framework.

That is precisely the advantage of technology-neutral regulation, as advocated by Luxembourg. It means regulating capabilities and risks rather than continually inventing new legal categories for every new architecture.

If additionally, we could get a real simplification of the existing set of rules, we would make a tremendous step forward in terms of competitiveness.

Let me put it with a bit of humour:

If the biblical Commandments would not have been handed down to Moses, but to the European Commission, we would probably not have 10, but 160 Commandments, completed by 400 recitals and 1000 footnotes.

If we want people and companies to abide by the laws, we need to get back to making laws that the average person will actually be able to read and understand more or less intuitively.

Isaac Asimov needed only three rules for his robots. I might even go one step further. Why not just apply article 1384 of the civil code to AI Agents?
"One is liable not only for damage caused by one's own actions, but also for damage caused by the actions of persons for whom one is responsible, or by things in one's custody."

But then again, this might shift too much of the actual policy making to jurisprudence. Not to mention that voters like to see politicians pass new legislation … as they might otherwise assume that they're not doing their job properly enough.

And then of course, there's the geographical dimension. An agent may be deployed in one jurisdiction, call a tool hosted in a second, and trigger a payment that settles in a third, within moments.

The topic is not entirely new. Take cloud computing. But with AI agents, it takes on a whole different dimension. No framework yet addresses that topic convincingly, and the result is an agent that may be compliant where it was deployed but not where its actions actually land. Hence, what we need is convergence at European and international level.

We also need convergence among regulators and supervisors. We can start in Luxembourg. One of the main discussion points with regard to the transposition of the AI Act in national law is the definition of the scope of each of the regulators. For Financial Institutions this means facing a matrix of supervisors.

Fortunately, the conclusion was not to set up an additional one on top of the existing ones. As, oddly, some AI industry players in the US have recently been calling for.

When it comes to agentic AI, zero risk is not an option. Finance has never operated according to a zero-risk principle.

This brings me to the 4 policy principles that should apply to agentic AI:

1) Proportionality

   The more consequential the action, the stronger the safeguards should be. An internal agent helping to organise documents should not necessarily have the same governance requirements as an autonomous system taking decisions that directly affect customers or move financial assets. A five-person Luxembourg fintech and a globally systemic financial institution should not automatically carry identical compliance burdens. This reasoning of course only works if this 5-person fintech is not, in reality, a 10,000 agents player that moves substantial capital daily. It is only a matter of time before we see the first hedge fund run entirely by Agents. The AI Act foresees proportionality in its sanctions, in particular as set out in article 99. We might have to push that logic further.

2) Human oversight — but what kind?

There's a large consensus that AI must remain "human-centred". According to the latest CSSF-BCL survey, 90% of reported AI use cases involve direct human intervention in critical decisions. A machine proposes something. A human approves it. That's what we call human in the loop. That works when the machine makes ten decisions per day. It becomes less convincing when an agentic system makes ten thousand micro-decisions per minute.

If a human simply clicks "approve" a thousand times, that would hardly be meaningful oversight. And there would not be any efficiency gains. So we may have to move from human in the loop to human on the loop. The human supervises the system. Defines its goals and boundaries.
Monitors its behaviour and intervenes when necessary.
That requires a different kind of control. And it requires something else: the controls cannot exist only inside the model. Every system that can be accessed by agents must be configured and safeguarded accordingly.

As autonomy increases, technical permissions, access rights, transaction limits, logging and independent monitoring become more important, not less.

3) Accountability must scale with autonomy.

The more autonomy a system has, the more clearly we must know who is responsible for deploying it, supervising it and defining the limits within which it operates. Not every person involved in an AI value chain has identical responsibilities.

The AI Act deliberately distinguishes providers from deployers and other actors. That distinction can change depending on what an institution actually does to a system.

For example, under Article 25 of the AI Act, a deployer of a high-risk AI system can become a provider in certain circumstances. The practical question when procuring agentic AI is therefore not only who is liable if it fails, but what obligations we assume by customizing and operating it. Under the AI Act, a bank remains responsible for governance, oversight and risk management obligations, even when its AI Agents are developed, operated or maintained by a third party. And then there is the end user, who carries a share of responsibility as well.

The tricky part is this:
If harm arises from an agent doing exactly what the person operating it instructed, that is a different question from harm arising from an agent acting on its own initiative.
Especially if the Agent autonomously uses tools not explicitly foreseen. This is what some call the "third party tool problem".
Therefore, our legal framework may need to distinguish liability for negligence from liability for genuine misconduct, rather than treating every failure identically.

And finally, there's the matter of possible conflicts of interest.
What if an AI agent decides to prioritize the best interests of their developers and deployers over those of their users or the user's clients?
Currently, there are no requirements or standardized mechanisms for addressing or even disclosing such conflicts of interest. This is not a new problem per se.
For example: Who is a private banking advisor ultimately loyal to, when recommending a given investment?
We have found solutions for this when the advisor is a human. We shall find them when he's an AI Agent as well.

4) Trust through transparency

Ultimately, the decisive factor for adoption in finance will not be technological capability. It will be trust.

This is also reflected in our national AI strategy. The strategy identifies trustworthy AI as a condition for sustainable adoption. People will not accept autonomous systems unless they believe those systems are reliable and secure. Especially if they are to handle their money.
This implies transparency, robustness, accountability, cybersecurity and data protection.
These elements must be embedded in AI systems, to become "trustworthy by design". And, I might add, by third party certification. Yet, the CSSF and BCL findings indicate that nearly half of reported AI use cases did not achieve good auditability and explainability.
With complex agentic AI systems this will become even more challenging.
Yet, without it, the human-on-the-loop concept cannot work. And I am not sure it could be solved by putting AI alone in charge of auditing and supervising AI.

Since August 2026, the AI Act Article 50 transparency duties apply and foresee that a customer dealing with a customer-facing agent in a Luxembourg institution is entitled to know that they are dealing with a machine.

Let me, once again, be slightly provocative. The measure is supposed to convey trust. But it might just do the opposite. It translates distrust in AI and Agents. Imagine if there was an analogue disclaimer when faced with inexperienced junior staff, or, if in the Seventies, you would have had to inform your customers that you did your calculations no longer by hand, but with one of those new PCs.

So, the real question is: do we see agents as potentially dangerous tools, or as actors in their own right?

Is Luxembourg ready?

AI adoption in the Luxembourg financial sector is steadily increasing. We have the ecosystem, the legal framework, the tax incentives and the infrastructure. Technically speaking, we're ready. But is that enough?

Let's face it: Agentic AI has the potential, even more so than LLMs, to radically change tasks and professions. Jobs will disappear. People will not take these changes lightly. Not in Luxembourg. Not anywhere else. But we have no choice. We have to adapt and see AI, and the agents, as colleagues, rather than competition.

AI literacy, training and continuous reskilling therefore needs to become a natural reflex. But not everyone needs to become an AI engineer. The AI Act has been rightly toned down in this regard. We also do not require someone to pass a test in electricity before allowing them to switch on a lightbulb. Obviously, it's a whole different story for those who develop and install them. So, again: proportionality is key.

Let me finish with one topic that I believe deserves particular attention, and which our AI Advisory board has identified as an additional opportunity for Luxembourg.

As AI agents are being deployed at scale, two questions become unavoidable: Who is this agent acting for? And is it allowed to do that? In other words: is there a verifiable chain of authority?

Once again: what seems to be a very new question is, in civil law terms, a very old one. It's a matter of mandate and representation. Yet, as agents are, to some extent, immaterial beings, the classic toolbox is insufficient. To come back to my earlier question: do we want to see agents as tools or as actors in their own right? In the latter case, wouldn't they need some form of legal personality, or, at the least, legal identity? Because autonomy should not lead to anonymity.

In practice, Agents already open accounts and do transactions, often without anyone having explicitly authorised them to do so. Hence, we need to ensure that the institution or person that deployed the agent and/or benefits from it can always be clearly identified. For this, we need a ledger of agentic IDs. The technological building blocks are beginning to exist.

The eIDAS 2 Regulation allows for legally recognised electronic evidence of attributes. A qualified trust service provider such as LuxTrust can attest to a mandate or a power of representation digitally, with the same legal effect as its paper equivalent. (Electronic Identification, Authentication, and Trust Services).

The Global Legal Entity Identifier Foundation (GLEIF) is now developing the verifiable LEI, which adds a cryptographically verifiable link between the entity and the person acting for it. In legal terms, think of it as combining a company-register extract, proof of identity and a power of attorney in a machine-readable credential. It is obvious how such a system could be useful in a world of AI Agents.

Providing the relevant legal certainty and infrastructure might give Luxembourg an additional edge with regard to Agentic AI.

Conclusion

For Luxembourg, agentic AI is more than a regulatory challenge. It is also a strategic opportunity. If we succeed in implementing the appropriate policies, create legal certainty, and encourage further innovation, Luxembourg can help turn what was once science fiction into trusted financial infrastructure.