



**PauseAI UK**  
**5 Brayford Square**  
**London E1 0SG**

**Sir Demis Hassabis**  
**Chief Executive Officer, Google DeepMind**  
**London, United Kingdom**

**28 July 2025**

Dear Sir Demis,

We write to express profound concern about Google DeepMind's failure to honour the Frontier AI Safety Commitments signed at the AI Seoul Summit in 2024. The release of Gemini 2.5 Pro without the transparency required by paragraph VIII of the commitments represents a troubling breach of trust with governments and the public.

At the AI Seoul Summit, Google explicitly committed to:

- Conducting safety tests "before deploying" AI models with input from "independent third-party evaluators" as appropriate.
- Providing "public transparency" into testing processes.
- Disclosing "how, if at all, external actors, such as governments... are involved in the process".

Yet when you released Gemini 2.5 Pro on 28 March, no safety evaluation report accompanied it. A month later, only a minimal "model card" appeared, lacking any substantive detail about external evaluations. Even when directly questioned by journalists, Google refused to confirm whether government agencies like the UK AI Security Institute participated in testing.

This is not a matter of semantics or technicalities. Labelling a publicly accessible model as "experimental" does not absolve Google of its safety obligations. When anyone on the internet can use a frontier AI system, it has been deployed in every meaningful sense.

You yourself have stated that AGI may arrive within five years. Leading AI researchers, such as Geoffrey Hinton and Yoshua Bengio, estimate a 10% or greater chance that advanced AI could cause human extinction. These are not distant hypotheticals but near-term possibilities requiring immediate, serious action.

We are particularly troubled that Google, having helped establish these safety standards, would be among the first to abandon them. This sets a dangerous precedent that undermines global efforts to develop AI safely. If industry leaders treat safety commitments as optional when convenient, how can we expect others to take them seriously?

**We therefore call on Google DeepMind to:**

1. **Establish clear definitions** of "deployment" that align with common understanding - when a model is publicly accessible, it is deployed.
2. **Publish a specific timeline** for when safety evaluation reports will be released for all future models.
3. **Clarify unambiguously**, for each model release, which government agencies and independent third-parties are involved in testing, and the exact timelines of their testing procedures.

The development of artificial general intelligence may be humanity's most consequential undertaking. It demands the highest standards of responsibility, transparency, and caution. Google's technical capabilities come with commensurate obligations to society.

We await your response and concrete actions to address these critical concerns.

Yours sincerely,