

Advancing Responsible AI: Hirundo's Unlearning Platform Cuts Bias by Nearly Half in State-of-the-Art Llama 4

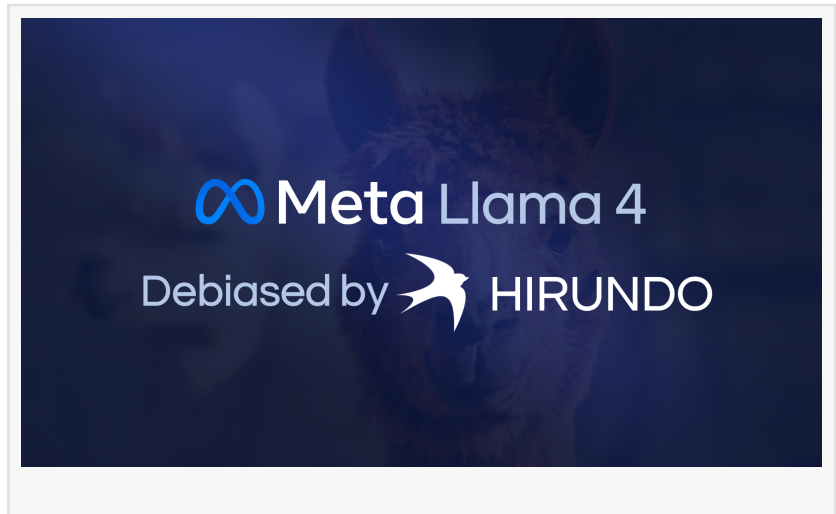
Hirundo, a machine unlearning startup, announced today a significant advancement in responsible AI by reducing biases in Meta's Llama 4 model by nearly half.

TEL AVIV, ISRAEL, April 10, 2025 /EINPresswire.com/ -- Hirundo, a pioneering machine unlearning startup, [announced today](#) a significant advancement in responsible AI by reducing biases in the newly released state-of-the-art Llama 4 (Scout) model by an impressive average of 44%. This achievement underscores Hirundo's unique ability to significantly enhance AI model fairness and safety through its proprietary machine unlearning platform, even in large-scale AI deployments.

Llama 4 (Scout), developed by Meta, is a 17-billion-parameter model utilizing a Mixture-of-Experts (MoE) architecture with 16 experts, totaling 109 billion parameters. Released earlier this week after long anticipation, it was quickly celebrated for its native multimodal capabilities, efficiently processing both text and images, and supporting an extensive context window of up to 10 million tokens - the largest among publicly released models. Given its recent release and promising capabilities, addressing inherent biases early is crucial for its safe adoption in sensitive applications within finance, healthcare, legal services, and beyond.

Machine unlearning is an emerging approach that enables targeted removal or suppression of undesired data or behaviors in AI models - such as bias, hallucination, or toxicity - without the need to retrain from scratch. In essence, it's about "making AI forget".

Leveraging its innovative machine unlearning platform, Hirundo successfully mitigated these biases without compromising model performance. This milestone builds on Hirundo's previous success with smaller models, such as DeepSeek-R1-Distill-Llama (8B parameters), highlighting the scalability and effectiveness of its approach across models of varying sizes and



complexities.

"Bias reduction is fundamental to the responsible adoption of advanced AI models," said Ben Luria, CEO of Hirundo. "Our work with Llama 4 demonstrates the robustness and scalability of our platform, reinforcing our commitment to helping organizations deploy safer, fairer AI solutions."

Hirundo's machine unlearning methodology extends beyond bias mitigation, effectively addressing other key AI behaviors such as hallucinations, adversarial vulnerabilities, and toxic outputs. Enterprises and data scientists can leverage Hirundo's customizable platform to efficiently adapt their AI models to evolving ethical standards and regulatory requirements.

"We encourage enterprises and AI professionals to explore the transformative capabilities of our machine unlearning platform," said Michael Leybovich, CTO of Hirundo. "We are dedicated to supporting organizations in achieving ethical, compliant, and trustworthy AI deployments."

Hirundo has made the debiased version of Llama 4 (Scout) publicly available on [Hugging Face](#).

For additional information about Hirundo's breakthrough work or to access the debiased Llama 4 (Scout) model, visit [their website](#) or contact the Hirundo team directly for tailored demonstrations and support.

About Hirundo:

Hirundo is dedicated to developing advanced machine unlearning solutions that empower organizations to deploy responsible, safe, and ethically-aligned AI models. Hirundo's proprietary technology efficiently mitigates harmful behaviors in AI, ensuring compliance, enhancing user trust, and facilitating reliable AI deployments across various industries.

Ben Luria

Hirundo

+972 50-480-7752

ben@hirundo.io

Visit us on social media:

[LinkedIn](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/802042028>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2025 Newsmatics Inc. All Right Reserved.