

Supplementary information. The online version of this article contains supplementary material available in Zenodo at <https://zenodo.org/records/18098196>. The archived material includes the source code for quantized LLM inference, evaluation, and figure/table generation, together with extended performance benchmarks for the quantized LLM on embedded platforms and additional semantic telemetry visualization traces.

For software traceability, the AI workflow used the archived software package *Bucle2D*, version 4.0, publicly deposited in Zenodo at <https://zenodo.org/records/18098196>, together with GPTQ-based quantization tools implemented through the official AutoGPTQ project. The structural engineering workflow was performed using ANSYS Workbench 2023 / Ansys Mechanical 2023 for topology optimization and finite-element analysis, and SOLIDWORKS 2023 for CAD geometry preparation.

Proprietary software was accessed under valid institutional or commercial licenses, whereas open-source software was used in accordance with the corresponding license terms. The manuscript and supplementary materials report only original outputs generated with these tools and do not redistribute proprietary software, binaries, protected graphical assets, or third-party model weights.

Model weights are subject to the respective licenses of their original providers and are therefore not redistributed. The repository provides scripts and configuration files to reproduce the evaluation using the referenced base models under their corresponding license terms.