

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
ARTIFICIAL INTELLIGENCE

# An effective strategy for reducing the size of LLaMA-based Language Models

Supervisor:  
Prof. Francesco Conti

Presented by:  
Angelo Galavotti

Co-supervisor:  
Luca Bompani

Sessione I  
Anno Accademico 2024/2025