

الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

المدرسة العليا للإعلام الآلي 8 ماي 1945 - سيدي بلعباس

Higher School of Computer Science

8 Mai 1945 - Sidi Bel Abbès



Thesis

To obtain the diploma of **Engineer's Degree**

Field of Study: **Computer Science**

Specialization: **Artificial Intelligence & Data Science**

Theme

Aligning Large Language Models at Inference-Time with Preference Vectors using Hypernetworks

Presented by

Dallaa Abdelsamad Taki-Eddine

Defended on: **July 4th, 2026**

In front of the jury composed of:

Mr. Belkacem Khaldi	President of the Jury
Mr. Souleyman Chaib	Supervisor
Mr. Zeyd Boukhers	Supervisor
Mr. Lingxiao Kong	Supervisor
Mrs. Rabab Bousmaha	Examiner

Academic Year: 2025/2026

Abstract

Modern [Large Language Model \(LLM\)](#)s deployed in real-world settings must simultaneously satisfy multiple, often conflicting, user objectives, such as helpfulness, safety, verbosity, and correctness. Standard fine-tuning collapses these trade-offs into a single fixed behaviour, leaving users without recourse when their needs deviate from the model’s learned prior. Multi-objective alignment methods address this by producing a family of behaviours parameterised by a *preference vector* on the probability simplex, so that inference-time steering becomes possible.

Existing preference-conditioned adapters inject the preference vector into the singular-value scaling factors of a frozen [Singular Value Decomposition \(SVD\)](#)-based [Low-Rank Adaptation \(LoRA\)](#) adapter. This formulation assumes that the optimal trade-off for a given preference vector is *context-independent*, an assumption that breaks down in practice. A coding query and a medical query may require very different weightings of helpfulness versus safety even under an identical preference vector.

This thesis proposes a lightweight hypernetwork that conditions the singular-value adapter on *both* the preference vector and the input prompt x , using Feature-wise Linear Modulation ([FiLM](#)) to fuse the two sources of context. The hypernetwork generates per-layer, per-matrix singular-value vectors in approximately 630 K trainable parameters, leaving the base model fully frozen. Training uses reference-free [Simple Preference Optimisation \(SimPO\)](#) with an IRPO-inspired NLL regularization term, Dirichlet-sampled preferences, and Tchebycheff scalarisation for winner determination, enabling dense coverage of the preference simplex while saving memory by not having to load a reference model or [Supervised Fine-Tuning \(SFT\)](#) warm-start.

All methods discussed in this thesis are evaluated on the HelpSteer2 dataset using Qwen2.5-0.5B-Instruct as the backbone, with hypervolume and linear utility as primary metrics.

Keywords: Large Language Models, Multi-Objective Alignment, Hypernetworks, Parameter Efficient Fine-Tuning, Pareto Optimality, Preference Vectors, Singular Value Decomposition, SimPO, IRPO

Résumé

Les modèles LLM déployés en situation réelle doivent satisfaire simultanément de multiples objectifs utilisateurs, souvent contradictoires, tels que l'utilité, la sécurité, la concision et l'exactitude. Le fine-tuning supervisé standard (SFT) réduit ces compromis à un comportement fixe unique, privant ainsi les utilisateurs de toute possibilité de modification lorsque leurs besoins s'écartent des connaissances a priori acquises par le modèle. Les méthodes d'alignement multi-objectifs remédient à ce problème en générant une famille de comportements paramétrés par un vecteur de préférence p sur le simplexe de probabilité, permettant ainsi un pilotage lors de l'inférence.

Les adaptateurs conditionnés par les préférences existants injectent p dans les facteurs d'échelle des valeurs singulières d'un adaptateur LoRA basé sur la décomposition en valeurs singulières (SVD) figé. Cette formulation suppose que le compromis optimal pour un vecteur de préférence donné est indépendant du contexte, une hypothèse qui se révèle erronée en pratique. Une requête de code et une requête médicale peuvent nécessiter des pondérations très différentes entre l'utilité et la sécurité, même avec un vecteur de préférences identique.

Cette thèse propose un hyperréseau léger qui conditionne l'adaptateur de valeurs singulières à la fois au vecteur de préférences et à la requête d'entrée x , en utilisant la modulation FiLM pour fusionner les deux sources de contexte. L'hyperréseau génère des vecteurs de valeurs singulières par couche et par matrice, avec environ 630 000 paramètres entraînaibles, le modèle de base restant entièrement figé.

L'entraînement utilise l'optimisation simple des préférences (SimPO) sans modèle de référence, avec des préférences échantillonnées à partir d'une distribution Dirichlet et une scalarisation de Tchebycheff pour la détermination de la meilleure réponse, permettant une couverture dense du simplexe de préférences sans modèle de référence ni initialisation par fine-tuning supervisé (SFT).

Toutes les méthodes discutées sont évaluées sur le jeu de données HelpSteer2 en utilisant Qwen2.5-0.5B-Instruct comme architecture de base, avec l'hypervolume et l'utilité linéaire comme métriques principales.

Mots-clés: LLM, Alignement multi-objectif, Hyperréseaux, Optimisation paramétrique efficace, Optimalité de Pareto, Vecteurs de préférence, Décomposition en valeurs singulières, SimPO, IRPO.

ملخص

يجب على نماذج LLM الحديثة المستخدمة في بيئات العالم الحقيقي أن تُلبّي أهدافاً متعددة للمستخدمين في آنٍ واحد، غالباً ما تكون متناقضة، مثل الفائدة، والأمان، والإسهاب، والصحة. يُختزل الـ Fine-Tuning Supervised (SFT) العادي هذه المفاضلات إلى سلوك ثابت واحد، مما يترك المستخدمين دون أي سبيل للرجوع إليه عندما تختلف احتياجاتهم عن المعلومات المسبقة المكتسبة من النموذج. تعالج أساليب الـ Multi-Objective Alignment هذه المشكلة من خلال إنتاج مجموعة من السلوكيات المُعلّمة بواسطة شعاع تفضيل على مُركب الاحتمالية، بحيث يصبح التوجيه في وقت inference ممكناً.

تقوم المُهايات الحالية المُشروطة بالتفضيل بإدخال شعاع التفضيل في عوامل قياس القيمة المفردة لمُهايات LoRA المُجمّد القائم على تحليل القيمة المفردة (SVD). يفترض هذا النموذج أن المفاضلة المثل لقيمة شعاع تفضيل مُعيّن مستقلة عن السياق، وهو افتراض ينهار عملياً. قد يتطلب سؤال طبي وسؤال برنجي أوزاناً مختلفة تماماً للفائدة مقابل السلامة حتى في ظل شعاع تفضيل متطابق.

تقترح هذه الأطروحة شبكة hypernetwork خفيفة الوزن تُقيّد مُهايات القيم المفردة على كلٍ من شعاع التفضيل ومُدخل x ، باستخدام (FiLM) لدمج مصدرين السياق.

تُولد الشبكة الفائقة متجهات قيم مفردة لكل طبقة ولكل مصفوفة في حوالي 630 ألف مُعامل قابل للتدريب مع بقاء النموذج الأساسي ثابتاً تماماً. يستخدم SimPO بدون مرجع مع تفضيلات مُختارة وفقاً لنموذج ديريشليه وتحويل تشيبيشيف لتحديد أفضل إجابة، مما يُتيح تغطية كثيفة لمصفوفة التفضيل البسيطة بدون نموذج مرجعي أو بدء تشغيل مُسبق باستخدام Fine-Tuning Supervised (SFT).

تم تقييم جميع الطرق على مجموعة بيانات HelpSteer2 باستخدام Qwen2.5-0.5B-Instruct كبنية أساسية، مع اعتبار الـ hypervolume والـ Utility Linear كمقاييس أساسية.

الكلمات المفتاحية: نماذج اللغة الكبيرة، محاذاة الأهداف المتعددة، الـ hypernetworks الضبط الدقيق الفعال للمعلّمت، أمثلة باريتو، متجهات التفضيل، تحليل القيم المفردة، SimPO، IRPO.