

IST/2026/BL118

Bolsa de Investigação para alunos matriculados em curso de Doutoramento na área científica de Engenharia eletrotécnica, electrónica e informática

Orientador Científico: Arlindo Manuel Limede de Oliveira (ist12282)

Coorientador(es): Rui Miguel Carrasqueiro Henriques (ist156846), INESC-ID.

Unidade Orgânica: Departamento de Engenharia Informática

Tema da Bolsa: Raciocínio Latente Recursivo em Modelos de Linguagem Baseados em Difusão

Duração Inicial da Bolsa: 12 meses

Duração Máxima Incluindo Renovações: 48 meses

Subsídio de Manutenção Mensal: 1359,64 €

Entidade Financiadora: Instituto Superior Técnico (IST)

Programa Operacional: Plano de Recuperação e Resiliência (PRR)

Objetivos

Esta tese investigará se a integração de mecanismos de refinamento latente recursivo em modelos de linguagem baseados em difusão pode melhorar a robustez e o raciocínio sistemático face a arquiteturas autoregressivas. Serão desenvolvidos e avaliados modelos que combinam denoising iterativo com supervisão profunda, incluindo comparações controladas e análises mecanísticas das dinâmicas de raciocínio, como retrocesso, convergência e modos de falha. O trabalho pretende fornecer evidência empírica e enquadramento teórico sobre o papel da computação iterativa em modelos de linguagem, bem como avanços práticos em eficiência e em arquiteturas escaláveis para raciocínio.

Plano de Trabalho

Ano 1 – Fundamentos e Modelos de Referência

O primeiro ano será dedicado a uma revisão aprofundada da literatura sobre modelos de linguagem autoregressivos e baseados em difusão, com foco nas capacidades de raciocínio e computação iterativa. Serão reproduzidos modelos autoregressivos de referência e estendidos com mecanismos de refinamento recursivo inspirados em trabalhos recentes. Será também estabelecido um conjunto de benchmarks para tarefas de raciocínio, juntamente com ferramentas iniciais de avaliação e análise.

Ano 2 – Desenvolvimento de Modelos Baseados em Difusão

O segundo ano centrar-se-á na implementação e treino de modelos de linguagem baseados em difusão em condições computacionais controladas. Serão integrados mecanismos de refinamento latente recursivo com supervisão profunda. Serão conduzidas experiências comparativas entre abordagens autoregressivas e baseadas em difusão, assegurando uma avaliação justa nos benchmarks definidos.

Ano 3 – Análise Mecanística e Optimização

O terceiro ano será dedicado ao estudo das dinâmicas internas de raciocínio em ambas as famílias de modelos, analisando comportamentos de convergência, modos de falha e o papel do refinamento iterativo. Serão realizados estudos de ablação controlados para isolar o impacto da recursividade, da arquitectura e das estratégias de treino. Serão também exploradas melhorias de eficiência, incluindo compressão e optimização de modelos.

Ano 4 – Escala, Validação e Escrita da Tese

O último ano focar-se-á na escalabilidade das abordagens mais promissoras e na realização de avaliações abrangentes face ao estado da arte. Os resultados serão consolidados em publicações científicas e na dissertação de doutoramento. A validação final, documentação e disseminação dos resultados concluirão o projecto.

Requisitos de Admissão

Experiência em modelos de linguagem autoregressivos e baseados em difusão. Conhecimentos dos fundamentos teóricos das arquiteturas de modelos de linguagem. O candidato deverá estar inscrito num programa doutoral adequado à área de trabalho.

Legislação e Regulamentação Aplicável

Lei n.º 40/2004, de 18 de agosto (Estatuto do Bolseiro de Investigação Científica), na sua redação atual; Regulamento de Bolsas de Investigação do IST, disponível em

https://drh.tecnico.ulisboa.pt/files/sites/45/despacho_8532_regulamento_bolsas.pdf.

Enquadramento, Carga de Trabalho e Horário

Local de Trabalho: INESC-ID

Regime de Prestação de Trabalho: Não aplicável.

Campus Principal: Alameda

Carga Média Semanal Indicativa: Não aplicável.

Horário de Prestação de Funções Indicativo: Não aplicável.

Metodologia de Avaliação do Concurso

Avaliação curricular ponderado a 100% numa escala de 100 valores com um mínimo de 50 valores para admissão.

Valor final mínimo para admissão de 50 valores.

Condições para a Realização dos Métodos de Seleção

40% - currículo académico e notas

40% - experiência na área

20% - publicações

Composição do Júri de Seleção

Presidente do Júri: Arlindo Manuel Limede de Oliveira (ist12282)

Vogais: Rui Miguel Carrasqueiro Henriques (ist156846), INESC-ID; Bruno Emanuel Da Graça Martins (ist24686), INESC-ID.

Em caso de impossibilidade do presidente do júri, este será substituído por um dos vogais efetivos.

Tramitação do Concurso

A apresentação de candidaturas é efetuada exclusivamente na [plataforma de admissões](#) do Instituto Superior Técnico em <https://fenix.tecnico.ulisboa.pt/fenixedu-admissions> e requer registo e validação de identidade dos candidatos.

As candidaturas só são formalizadas quando o formulário disponível na plataforma é devidamente preenchido, submetido e lacrado sem erros de validação. A documentação obrigatória a ser anexada no formulário para esta bolsa inclui os seguintes documentos:

Curriculum Vitae

Certificado de Habilitações (ou compromisso de honra caso não tenha ainda terminado o curso)

Comprovativo de Inscrição/Matrícula

Carta de Motivação

Os prazos para a submissão das candidaturas devem ser consultados na mesma plataforma de admissões.

Os resultados do concurso serão disponibilizados na mesma plataforma de admissões.



IST/2026/BL118**Research Scholarship of Research
for students registered in a Doctoral Programme
for the scientific area of Electrical, electronic and information engineering**

Scientific Advisor: Arlindo Manuel Limede de Oliveira (ist12282)

Co-advisor(s): Rui Miguel Carrasqueiro Henriques (ist156846), INESC-ID.

Organic Unit: Department of Computer Science and Engineering

Scholarship Theme: Recursive Latent Reasoning in Diffusion Language Models

Duration: 12 months

Maximum Duration Including Renewals: 48 months

Monthly Maintenance Allowance: €1,359.64

Funding Entity: Instituto Superior Técnico (IST)

Operational Programme: Recovery and Resilience Plan (RRP)

Objectives

This dissertation will investigate whether integrating recursive latent refinement mechanisms into diffusion-based language models can improve robustness and systematic reasoning compared to autoregressive architectures. It develops and evaluates models that combine iterative denoising with deep supervision, while performing controlled comparisons and mechanistic analyses of reasoning dynamics such as backtracking, convergence, and failure modes. The work aims to provide both empirical evidence and theoretical insight into the role of iterative computation in language models, alongside practical advances in efficiency and scalable reasoning architectures.

Work Plan**Year 1 – Foundations and Baselines**

The first year will focus on a comprehensive literature review covering autoregressive and diffusion-based language models, with emphasis on reasoning capabilities and iterative computation. Baseline autoregressive models (e.g., GPT-style) will be reproduced and extended with recursive refinement mechanisms inspired by recent work. A benchmark suite for reasoning tasks will be established, together with initial evaluation and analysis tools.

Year 2 – Diffusion-Based Model Development

The second year will focus on implementing and training diffusion-based language models under controlled computational settings. Recursive latent refinement mechanisms with deep supervision will be integrated into these models. Comparative experiments between autoregressive and diffusion-based approaches will be conducted, ensuring fair evaluation across benchmarks.

Year 3 – Mechanistic Analysis and Optimization

The third year will explore the internal dynamics of reasoning in both model families, analysing convergence behaviour,

failure modes, and the role of iterative refinement. Controlled ablation studies will be carried out to isolate the effects of recursion, architecture, and training strategies. Efficiency improvements will be pursued, including model compression and optimization techniques.

Year 4 – Scaling, Validation, and Thesis Writing

The final year will focus on scaling the most promising approaches and performing comprehensive evaluations against state-of-the-art models. The results will be consolidated into scientific publications and the doctoral dissertation. Final validation, documentation, and dissemination of outcomes will complete the project.

Admission Requirements

Experience with autoregressive and diffusion-based language models. Knowledge of the theoretical foundations of language model architectures. The candidate must be enrolled in an appropriate doctoral program.

Applicable Laws and Regulations

Law No. 40/2004, of 18 August (Statute of Scientific Research Fellow), in its current wording; FCT Regulation for Research Fellowships, available at

https://drh.tecnico.ulisboa.pt/files/sites/45/despacho_8532_regulamento_bolsas.pdf.

Context, Workload and Schedule

Workplace: INESC-ID

Work Model: Not applicable.

Main Campus: Alameda

Expected Average Weekly Workload: Not applicable.

Expected Schedule for Activities and Functions: Not applicable.

Contest Evaluation Method(s)

Curricular evaluation weighted to 100% on a scale of 100 points with a minimum of 50 points needed for admission.

The minimum final grade needed for admission is 50 points.

Conditions for the Contest Evaluation

40% - academic curriculum and grades

40% - previous experience

20% - publications

Composition of the Selection Jury

Jury President: Arlindo Manuel Limede de Oliveira (ist12282)

Jury Members: Rui Miguel Carrasqueiro Henriques (ist156846), INESC-ID; Bruno Emanuel Da Graça Martins (ist24686), INESC-ID.

In case the president of the jury is unable to preside, they will be replaced by one of the jury members.

Contest Procedure

Applications must be exclusively submitted on the [admissions platform](https://fenix.tecnico.ulisboa.pt/fenixedu-admissions) of the [Instituto Superior Técnico](https://fenix.tecnico.ulisboa.pt/fenixedu-admissions) at <https://fenix.tecnico.ulisboa.pt/fenixedu-admissions> and requires registration and validation of the candidate's identity.

Applications are only accepted when the form available in the platform is correctly filled, submitted and locked without any validation errors. The mandatory documentation to submit in the scholarship application includes:

Curriculum Vitae

Proof of Qualifications (or declaration of honor in case you do not yet have the certificate)

Proof of Registration/Enrolment

Motivation Letter

The application submission deadlines can be viewed in the admissions platform.

The results of the contest will be made available in the same admissions platform.

