

## **IST/2026/BL108**

### **Bolsa de Investigação para alunos matriculados em curso de Doutoramento na área científica de Ciências da computação e da informação**

**Orientador Científico:** Alexandre Paulo Lourenço Francisco (ist14152)

**Coorientador(es):** Cátia Raquel Jesus Vaz (ist45098), INESC-ID Lisboa / ISEL, IPL.

**Unidade Orgânica:** Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento em Lisboa

**Tema da Bolsa:** Desenvolvimento de algoritmos baseados em modelos generativos e análise topológica de dados para análises filogenéticas em larga escala

**Duração Inicial da Bolsa:** 12 meses

**Duração Máxima Incluindo Renovações:** 36 meses

**Subsídio de Manutenção Mensal:** 1359,64 €

**Entidade Financiadora:** Instituto Superior Técnico (IST)

**Programa Operacional:** Plano de Recuperação e Resiliência (PRR)

### **Objetivos**

A reconstrução filogenética é um problema fundamental na biologia computacional e na bioinformática onde, dado um conjunto de sequências de organismos descendentes de um ancestral comum, o objetivo é reconstruir a árvore (binária) que descreve a evolução a partir deste último. Métodos de última geração baseados em Máxima Verossimilhança e Inferência Bayesiana, embora eficazes, compreendem um custo computacional elevado, o que limita sua aplicabilidade em grandes conjuntos de dados. Por outro lado, métodos combinatórios são mais escaláveis, porém menos eficazes. Abordagens de Deep Learning (aprendizagem profunda) foram consideradas recentemente para abordar este problema e, embora promissoras – com precisão comparável à da Máxima Verossimilhança –, os resultados ainda são limitados, nomeadamente se o objetivo for inferir padrões filogenéticos para grandes conjuntos de dados. Esta bolsa visa apoiar a investigação a nível de doutoramento sobre este problema, explorando novos métodos que combinam modelos generativos e estocásticos (ex: GFlowNet), redes neuronais para grafos (GNNs) e análise de dados topológicos (TDA), que demonstraram melhorar significativamente o desempenho preditivo em diversas tarefas noutras áreas. Os dados considerados incluem alinhamentos múltiplos de sequências e árvores evolutivas, para os quais representações adequadas também devem ser estudadas neste contexto.

### **Plano de Trabalho**

O plano de trabalhos a realizar consistirá em: (1) realizar o levantamento do estado da arte sobre a combinação de modelos generativos, redes neuronais para grafos (GNNs) e análise de dados topológicos (TDA), bem como de trabalhos recentes sobre abordagens baseadas em deep learning para a reconstrução filogenética; (2) estudar e investigar representações adequadas de alinhamentos múltiplos de sequências, dados de tipagem e árvores evolutivas para processamento posterior; (3) identificar e investigar de que forma os modelos generativos, GNNs e TDA podem melhorar os métodos de reconstrução de padrões filogenéticos, incluindo a superação do paradigma das árvores e a identificação

de padrões reticulados; (4) realizar avaliações experimentais comparativas de métodos existentes, bem como das abordagens propostas; (5) desenvolver implementações e ferramentas escaláveis com base nos resultados obtidos, seguindo uma abordagem baseada em engenharia de algoritmos; (6) redigir relatórios técnicos e artigos científicos sobre as conclusões e resultados obtidos.

Cronograma:

- Ano 1: Tarefas 1, 2, 3 e 6. Entrega de um levantamento bibliográfico (survey) e de um artigo a ser submetido a uma revista Q1/Q2 relacionado às tarefas 2 e 3.
- Ano 2: Tarefas 4 e 5; continuação das tarefas 3 e 6. Entrega de, pelo menos, um artigo a ser submetido a uma revista Q1/Q2 relacionado às tarefas 3 e 4.
- Ano 3: Continuação das tarefas 5 e 6. Entrega de um artigo a ser submetido a uma revista Q1/Q2 relacionado à tarefa 5, lançamento da ferramenta desenvolvida, e submissão da tese.

## Requisitos de Admissão

Os candidatos devem possuir licenciatura (BSc) e mestrado (MSc) em Engenharia Informática e de Computadores, Matemática Aplicada ou áreas afins; e estar devidamente registados e matriculados num doutoramento (PhD) em Engenharia Informática e de Computadores, ou áreas afins. Os candidatos devem ter experiência em machine learning (aprendizagem automática) e, em particular, em metodologias e tecnologias de deep learning (aprendizagem profunda), incluindo uma sólida base matemática. Devem possuir boas competências de programação, bem como boas capacidades de comunicação. Candidatos com interesse em biologia computacional e bioinformática serão valorizados.

## 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](https://drh.tecnico.ulisboa.pt/files/sites/45/despacho_8532_regulamento_bolsas.pdf).

## Enquadramento, Carga de Trabalho e Horário

**Local de Trabalho:** INESC-ID Lisboa, Rua Alves Redol 9, 1000-029 Lisboa, Portugal

**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 35% numa escala de 20 valores com um mínimo de 10 valores para admissão.

Entrevista individual ponderado a 30% numa escala de 20 valores com um mínimo de 10 valores para admissão.

Demonstrated experience in accordance admission criteria ponderado a 35% numa escala de 20 valores com um mínimo de 10 valores para admissão.

Valor final mínimo para admissão de 10 valores.

## Condições para a Realização dos Métodos de Seleção

Avaliação curricular ponderada a 35% numa escala de 20 valores com um mínimo de 10 valores para admissão. Análise da experiência demonstrada de acordo com os critérios de admissão ponderada a 35% com um mínimo de 10 valores para admissão. Entrevista individual ponderado a 30% numa escala de 20 valores com um mínimo de 10 valores para admissão.

## Composição do Júri de Seleção

**Presidente do Júri:** Alexandre Paulo Lourenço Francisco (ist14152)

**Vogais:** Arlindo Manuel Limede de Oliveira (ist12282), INESC-ID / IST, ULisboa; Cátia Raquel Jesus Vaz (ist45098), INESC-ID / ISEL, IPL.

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](https://fenix.tecnico.ulisboa.pt/fenixedu-admissions) do [Instituto Superior Técnico](https://fenix.tecnico.ulisboa.pt/fenixedu-admissions) 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.



Financiado pela  
União Europeia  
NextGenerationEU

**IST/2026/BL108****Research Scholarship of Research  
for students registered in a Doctoral Programme  
for the scientific area of Computer and information sciences**

**Scientific Advisor:** Alexandre Paulo Lourenço Francisco (ist14152)

**Co-advisor(s):** Cátia Raquel Jesus Vaz (ist45098), INESC-ID Lisboa / ISEL, IPL.

**Organic Unit:** Institute of Systems and Computer Engineering - Research and Development in Lisbon

**Scholarship Theme:** Development of generative model and topological data analysis based algorithms for large scale phylogenetic analyses

**Duration:** 12 months

**Maximum Duration Including Renewals:** 36 months

**Monthly Maintenance Allowance:** €1,359.64

**Funding Entity:** Instituto Superior Técnico (IST)

**Operational Programme:** Recovery and Resilience Plan (RRP)

**Objectives**

Phylogenetic reconstruction is a fundamental problem in computational biology and bioinformatics, where given a set of sequences from organisms descending from a common ancestor, the objective is to reconstruct the (binary) tree describing their evolution from the latter. State-of-the-art methods based on Maximum likelihood and Bayesian inference, although effective, have a high computational cost, which limits their usability on large datasets. Combinatorial methods on the other hand are more scalable, but less effective. Deep learning approaches have been recently considered to approach this problem and, although promising with its accuracy matching that of a Maximum likelihood, results are still limited, namely if we are aiming to infer the phylogenetic patterns for large datasets. This scholarship aims to support then PhD level research on this problem, exploiting novel methods combining generative and stochastic models (e.g., GFlowNet), graph neural networks (GNNs) and topological data analysis (TDA), which have shown to significantly improve predictive performance in several tasks in other areas. The data we are considering include multiple sequence alignments and evolutionary trees, for which suitable representations must also be studied in this context.

**Work Plan**

The work plan to be carried out will consist on: (1) survey recent work on the combination of generative models, graph neural networks (GNNs) and topological data analysis (TDA), as well as recent work on deep learning based approaches to phylogenetic reconstruction; (2) study and research suitable multiple sequence alignments, typing data and evolutionary trees representations for further processing; (3) identify and research how generative models, GNNs and TDA can improve methods for phylogenetic patterns reconstruction, including going beyond trees and identify reticulate patterns; (4) conduct experimental comparative evaluations of existing methods as well as of proposed approaches; (5) develop scalable implementations and tools based on the obtained results and following an algorithm engineering based

approach; (6) write technical reports and papers regarding findings and results.

Timeline:

- Year 1: Tasks 1, 2, 3 and 6. Delivery of a survey and a paper to be submitted to a Q1/Q2 journal related to tasks 2 and 3.

- Year 2: Tasks 4 and 5; continuation of tasks 3 and 6. Delivery of at least a paper to be submitted to a Q1/Q2 journal related to tasks 3 and 4.

- Year 3: Continuation of tasks 5 and 6. Delivery of a paper to be submitted to a Q1/Q2 journal related to task 5, release of the developed tool, and submission of the thesis.

## **Admission Requirements**

Candidates should have a BSc and an MSc in Computer Science and Engineering, Applied Mathematics, or similar; and be registered and enrolled in a PhD in Computer Science and Engineering, or similar. Candidates should have experience with machine learning, and in particular with deep learning methodologies and technologies, including a strong mathematical background. Candidates should have good programming skills as well as good communication skills. Candidates with interest in computational biology and bioinformatics will be valued.

## **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](https://drh.tecnico.ulisboa.pt/files/sites/45/despacho_8532_regulamento_bolsas.pdf).

## **Context, Workload and Schedule**

**Workplace:** INESC-ID Lisboa, Rua Alves Redol 9, 1000-029 Lisboa, Portugal

**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 35% on a scale of 20 points with a minimum of 10 points needed for admission.

Individual interview weighted to 30% on a scale of 20 points with a minimum of 10 points needed for admission.

Demonstrated experience in accordance admission criteria weighted to 35% on a scale of 20 points with a minimum of 10 points needed for admission.

The minimum final grade needed for admission is 10 points.

## **Conditions for the Contest Evaluation**

Curricular evaluation weighted to 35% on a scale of 20 points with a minimum of 10 points needed for admission.

Analysis of demonstrated experience in accordance admission criteria weighted 35% with a minimum of 10 points for admission. Individual interview weighted to 30% on a scale of 20 points with a minimum of 10 points needed for admission.

## Composition of the Selection Jury

**Jury President:** Alexandre Paulo Lourenço Francisco (ist14152)

**Jury Members:** Arlindo Manuel Limede de Oliveira (ist12282), INESC-ID / IST, ULisboa; Cátia Raquel Jesus Vaz (ist45098), INESC-ID / ISEL, IPL.

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.

