# PhD (+internship) job offer: combinatorial optimization on GPUs for decision-focused learning

## Subject

Decision-Focused Learning (DFL) [1] is an emerging field that bridges the gap between constrained optimization and machine learning. When combinatorial algorithms are used as layers inside a neural network, the decision pipeline can adapt to specific input data, thus scaling to larger and more intricate problems (e.g. with real-time or stochastic inputs). This new paradigm unlocks numerous practical applications to transportation and logistics, as shown by previous research from our group [2], [3], [4].

Unfortunately, DFL remains computationally expensive. The main culprits are calls to third-party optimization solvers, running on CPUs in a way that is difficult to parallelize. We are thus looking for a PhD student to design and implement GPU-friendly discrete optimization algorithms, which may be used to speed up data-driven decision pipelines. These algorithms must be tailored for modern parallel hardware [5], while taking advantage of batching and amortized preprocessing to handle many slight variants of the same underlying problem during training. Alongside methodological innovations, a major focus will be put on developing cross-platform open-source software, preferably (but not exclusively) in the Julia programming language [6], [7]. Analyzing the theoretical properties of the suggested algorithms will also be of interest, especially with respect to the DFL context, where backpropagation does not require high precision nor perfectly optimal solutions.

The emphasis will not be on generic solvers targeting a single large Mixed Integer Linear Program (MILP), unlike e.g. [8]. Instead, problem-specific combinatorial approaches will be considered and adapted to tackle multiple instances at once with different objectives but similar constraints:

1. Dynamic programming, as a warm-up technique to experiment with GPU-native DFL on simple problems like the knapsack.
2. Decomposition methods (Dantzig-Wolfe, Benders, Lagrangian relaxation), where additional parallelism can be leveraged across subproblems.
3. Local search (meta)heuristics, where additional parallelism can be leveraged across trajectories.
4. Graph algorithms, in collaboration with a postdoc from the team.

## Context

The position is offered by École nationale des ponts et chaussées (ENPC), which is part of Institut Polytechnique de Paris (IP Paris). It is hosted by the applied mathematics laboratory (CERMICS), in close proximity to the transportation laboratory (LVMT). Collaborations with foreign research teams working on similar topics are already underway.

The PhD student will be registered at the IP Paris doctoral school, and subject to all its rules.

## Practical details

- Supervisors: Dr. Guillaume Dalle (LVMT) & Prof. Axel Parmentier (CERMICS).
- Duration: 4-6 months (Master's internship) + 3 years (PhD).
- Starting date: negotiable, ideally early spring 2026.
- Salary: during the PhD, around 1800€ per month before taxes.
- Location: CERMICS, 6-8 Avenue Blaise Pascal, 77420 Champs-sur-Marne, France.

If awarded, the position will be fully funded, including salary, computer equipment and participation in conferences. Involvement in student mentorship and teaching will be encouraged, albeit not mandatory.

## How to apply

The prospective candidate must have obtained a master's degree (or be nearing completion of a master's degree) in either mathematics or computer science, with a specialization in at least one of the following subfields: machine learning, optimization, operations research, numerical algorithms, high-performance computing.

To apply, send an email to guillaume.dalle@enpc.fr and axel.parmentier@enpc.fr with the following documents:

- CV.
- Grade transcript from your undergraduate and graduate degrees.
- Cover letter (1 page max), tailored to the current offer.
- Sample of your scientific writing, like a research internship report.
- Sample of your software writing, like a coding project for a class you took (the language doesn't matter).
- Optional: a reference letter, e.g. from a previous supervisor.

The email subject must start with "[phd-gpu]". Incomplete applications will not be considered.

## Bibliography

[1] J. Mandi *et al.*, "Decision-Focused Learning: Foundations, State of the Art, Benchmark and Future Opportunities," *Journal of Artificial Intelligence Research*, vol. 80, pp. 1623–1701, Aug. 2024, doi: 10.1613/jair.1.15320.

[2] G. Dalle, L. Baty, L. Bouvier, and A. Parmentier, "Learning with Combinatorial Optimization Layers: A Probabilistic Approach." [Online]. Available: http://arxiv.org/abs/2207.13513

[3] L. Baty, K. Jungel, P. S. Klein, A. Parmentier, and M. Schiffer, "Combinatorial Optimization-Enriched Machine Learning to Solve the Dynamic Vehicle Routing Problem with Time Windows," *Transportation Science*, vol. 58, no. 4, pp. 708–725, July 2024, doi: 10.1287/trsc.2023.0107.

[4] T. Greif, L. Bouvier, C. M. Flath, A. Parmentier, S. U. K. Rohmer, and T. Vidal, "Combinatorial Optimization and Machine Learning for Dynamic Inventory Routing." [Online]. Available: http://arxiv.org/abs/2402.04463

[5] W.-m. W. Hwu, D. B. Kirk, and I. E. Hajj, *Programming Massively Parallel Processors: A Hands-on Approach.* Morgan Kaufmann, 2022.

[6] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, Jan. 2017, doi: 10.1137/141000671.

[7] T. Besard, C. Foket, and B. De Sutter, "Effective Extensible Programming: Unleashing Julia on GPUs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 4, pp. 827–841, Apr. 2019, doi: 10.1109/TPDS.2018.2872064.

[8] K. Perumalla and M. Alam, "Design Considerations for GPU-based Mixed Integer Programming on Parallel Computing Platforms," in ICPP Workshops '21. New York, NY, USA: Association for Computing Machinery, Sept. 2021, pp. 1–7. doi: 10.1145/3458744.3473366.