The main goal of this project (henceforth referred to as N3PDF) is to redefine the standards of the determination of the parton distribution functions (PDFs) that encode the structure of the proton, through modern machine learning (ML) techniques, with the goal of achieving percent or sub-percent accuracy. As explained in the Description of the Action (DoA), N3PDF is the R&D department of the NNPDF collaboration, which produces the most widely used PDF determinations. Specifically, N3PDF has focused on the development of methodologies and tools for PDF determination that are either too speculative and risky, or that require too long and dedicated a development time to be achievable by the NNPDF collaboration itself.

This goal is articulated into three main sub-headings, corresponding to the work packages (WP) of the DoA. For each WP the goal of arriving at results that have subsequently been adopted by NNPDF has been achieved. Consequently, the results of N3PDF have a dual nature: (1) tools and methodologies developed within N3PDF; (2) PDF determinations produced jointly with the NNPDF collaboration that exploit these tools. The tools and methodology are the unique contribution of N3PDF. These are now part of the common knowledge of the whole high-energy physics community (and beyond, see below), as N3PDF has adopted the FAIR principle and made all of its results fully public as open-access code. The NNPDF results based on them are the first application and showcase of these results.

We now discuss, for each WP, the main goals, the N3PDF results that have achieved them, and the NNPDF applications that they have led to.

WP1. The main goal of this WP, and of the whole project, is (see the DoA), "a fully automated suite of tools for PDF determination". This has been fully achieved: a machine-learning-based methodology for PDF determination first suggested in [9], which fully automatizes the choice of methodology through a K-folded hyperoptimization procedure, and has been implemented through a full restructuring, optimization and streamlining of the NNPDF code [54]. This methodology is referred to (and will henceforth be called) n3fit by NNPDF, reflecting its N3PDF origin. A number of ancillary results have enabled this, including theory results on PDF positivity [38], methodological developments such as feature scaling of neural networks [55], testing tools such as so-called "future tests" [36] and numerical optimization and hardware acceleration methodologies [18]. All are available as open-access code from public repositories, with extensive on-line documentation [54].
This has enabled a first PDF determination that reaches percent accuracies [71]. Such unprecedented accuracy is enabling precision phenomenology: in particular (together with some crucial results of WP2, to be discussed below) a first determination of the so-called "intrinsic charm" component of the proton [66]. This has also elicited interest in the popular press [D99].

WP2 The main goal is the construction of approximate N3LO PDFs and the associated determination of theory uncertainties on PDFs, leading to the construction of PDF sets with so-called missing higher-order uncertainties. This has also been fully achieved, though through a somewhat different route than envisaged in the DoA. It has required the construction of a fully novel suite of tools for PDF evolution and for the computation of deep-inelastic scattering processes: available tools did not offer the required functionalities both in terms of available options and in terms of producing results that can be used for the massive, fast computations needed for PDF determination and more generally for collider phenomenology. Specifically, two new codes: EKO [59,OD25] and YADISM [OD23]. These have been crucial ingredients in the determination of the intrinsic charm of the proton mentioned above. They are now part of the theory pipeline to be used for future NNPDF determinations. This in turn has required the development of a new interface that enables a unified computation of theory predictions for a variety of processes [68]. This will allow for future PDF determinations the replacement of the so-called K-factor approximation with exact computations. The format is also being adopted as a common interface by widely used codes that are used by the collider physics community, such as MadGraph and MATRIX.
This set of results is enabling the construction of a first NNPDF PDF set with theory uncertainties, as well as a first PDF set including state-of-the-art N3LO results. These PDF sets have been presented for the first time at the DIS23 conference https://bit.ly/3nuim0V .

WP3 The main goal is the construction of PDF sets that are as close as possible to what is actually being measured. Whereas in the DoA the emphasis was mostly on resummation and interfacing to Monte Carlo and parton showering, it has gradually emerged that for the needs of precision collider physics, an implementation of electroweak corrections would be rather more important, and efforts were concentrated on this task. QED effects were included in the previous generation of PDFs at an early stage [3]. With the availability of the high-precision PDFs from WP1, and the new theory pipeline of WP2, the main task has become the integration of electroweak corrections in this context. This task has been fully achieved: QED evolution and the photon PDF have been included in the EKO code discussed above, and a grid-based interface that allows for the fast inclusion of electroweak effects has been developed [57]. Indeed, the new theory pipeline discussed above is expressly designed in such a way that electroweak corrections can be obtained and included from available codes. A first set of high-precision PDFs including a photon PDF and QED evolution is being constructed by the NNPDF collaboration. Legenda: [x]: Publications; [ODx]: Open Data; [Dx]: Dissemination

1.2 If applicable: novel methodologies, and/or inter-disciplinary developments, and/or knowledge and technology transfer.

The N3PDF project combines high-energy physics goals and techniques with machine learning methodologies (ML) and statistics and data analysis tools, and is as such interdisciplinary in nature. Consequently, the set of tools and methodologies that have been its main deliverables are relevant in an increasingly wide hierarchy of contexts, ranging from the specifics of PDF determination to the very general context of numerical methods and data analysis techniques. These tools and methodologies are all made public in the form of open-source computer code.
We mention the most relevant ones under five main headings:
1-PDF determination:
* The main NNPDF code [OD27], which embodies the n3fit methodology, and includes the whole suite of tools for PDF determination. This is supplemented by a number of ancillary utilities:
* Evolutionary KERAS [OD9], for genetic minimization of PDF loss functions
* Pycompressor [OD17], for compression of Monte Carlo representation of PDFs
* Ganpdfs [OD16], for the generation of artificial PDF replicas
* Fiatlux [OD15], for the construction of photon PDFs
* Pdfflow [OD13], for porting PDFs on GPU hardware
* qPDFs [45]: a first PDF determination on a quantum computer
2-QCD:
* EKO [OD25], a new code for QCD evolution and kernel evolution library
* YADISM [OD23], a new deep-inelastic scattering module
Both of these introduce a number of functionalities that are not present in existing codes, and provide output in the form of a library of grid-based results that can be re-used without need for repeated computation
3-Collider physics:
* Pineappl [OD19], for the fast inclusion of electroweak corrections to hadronic processes
* Pineline [68], a common interface and standard between different codes for the computation of hadronic processes
Both these utilities aim at giving a common fast interfacing format for high-energy physics computations, of which PDF determination is merely a first application. For instance, the interface to Pineappl has been recently added to the xFitter QCD framework, see https://bit.ly/3pfJpx .
* madflow [OD20], a framework for Monte Carlo simulations of particle physics processes that exploits hardware acceleration
4-Machine learning
* The n3fit methodology [71], embodied in the current NNPDF code [D27]. This includes a number of novel applications of ML techniques, in particular for the accurate estimation of uncertainties in regression that can serve as a model for a variety of ML applications that go beyond PDFs.
* The future testing methodology [36], a novel procedure for the validation of ML tools.
5-Numerical methods and quantum computing
* Vegasflow [OD26], a code for hardware acceleration of numerical integration
* Qibo [OD21], an open-source operating system for quantum hardware
These are in part spin-offs of the N3PDF project, and they go well beyond it in different directions and with different scope.

1.3 Indicate what you would consider to be the (up to) five most significant achievements in your project (e.g. the five most important scientific publications and other research outputs, patents, interactions with stakeholders such as industry or policy makers, media reports or events, etc.).

The main achievement of the project (WP1) is surely the n3fit methodology, embodied in the NNPDF code [54], and the first PDF set [71] obtained using it.
The first showcase of this methodology is the discovery of evidence for an intrinsic charm component of the proton [66], which also had extensive coverage in the popular press.
Two important technical achievements (WP2 and WP3) are the EKO evolution code [59] and the Pineline framework and interface [68]. Currently, these are making possible NNPDF sets with missing higher-order uncertainties and with QED and then electroweak corrections. Moreover, the former has the potential of becoming the reference for QCD evolution, and the latter of providing a common interface for public codes for the computation of hadronic processes.

1.4 Would you consider any of these significant achievements as breakthroughs or as advancing a research field significantly beyond the state of the art? Were any of these unplanned/ unexpected? Give a brief explanation.

The n3fit methodology is a breakthrough on numerous counts:
-it completely revolutionizes the determination of the structure of the proton
-it is one of the first (if not the first) applications of machine learning in theoretical high-energy physics that goes beyond the proof-of-concept stage and actually substantially raises existing standards of reliability, accuracy, and precision
-it is one of the few applications of machine learning methods that include an accurate, detailed, and reliable estimate of uncertainties, validated through extensive testing and benchmarking
-it sets new standards in terms of open access and FAIR - it is in fact the only PDF determination software and one of the very few in high-energy physics that is fully open access with complete documentation
-it provides a number of utilities that give a common language for a wide span of collider physics applications, such as the Pineline interface and the EKO code
-it provides a testing ground and learning platform for machine learning, data analysis, and statistics applications to high energy physics, that will soon be exploited in a first advanced training school, see https://bit.ly/3LUqrFv .

1.5 Describe how the ERC grant has enabled you to establish or consolidate your research group (for Starting and Consolidator Grants).