The data-driven future of high-energy-density physics

https://doi.org/10.1038/s41586-021-03382-w

Received: 24 June 2020

Accepted: 22 February 2021

Published online: 19 May 2021

Check for updates

Peter W. Hatfield[1✉], Jim A. Gaffney[2✉], Gemma J. Anderson[2✉], Suzanne Ali[2], Luca Antonelli[3], Suzan Başeğmez du Pree[4], Jonathan Citrin[5], Marta Fajardo[6], Patrick Knapp[7], Brendan Kettle[8], Bogdan Kustowski[2], Michael J. MacDonald[2], Derek Mariscal[2], Madison E. Martin[2], Taisuke Nagayama[7], Charlotte A. J. Palmer[9], J. Luc Peterson[2], Steven Rose[1,8], J J Ruby[10], Carl Shneider[11], Matt J. V. Streeter[8], Will Trickey[3] & Ben Williams[12]

High-energy-density physics is the field of physics concerned with studying matter at extremely high temperatures and densities. Such conditions produce highly nonlinear plasmas, in which several phenomena that can normally be treated independently of one another become strongly coupled. The study of these plasmas is important for our understanding of astrophysics, nuclear fusion and fundamental physics—however, the nonlinearities and strong couplings present in these extreme physical systems makes them very difficult to understand theoretically or to optimize experimentally. Here we argue that machine learning models and data-driven methods are in the process of reshaping our exploration of these extreme systems that have hitherto proved far too nonlinear for human researchers. From a fundamental perspective, our understanding can be improved by the way in which machine learning models can rapidly discover complex interactions in large datasets. From a practical point of view, the newest generation of extreme physics facilities can perform experiments multiple times a second (as opposed to approximately daily), thus moving away from human-based control towards automatic control based on real-time interpretation of diagnostic data and updates of the physics model. To make the most of these emerging opportunities, we suggest proposals for the community in terms of research design, training, best practice and support for synthetic diagnostics and data analysis.

'Set the controls for the heart of the Sun' encouraged a 2004 paper[1] (riffing on the 1968 Pink Floyd song), describing the bright future of using Earth-based experiments to create conditions similar to those inside the Sun in the laboratory. Seventeen years later, substantial advances have been made in this research programme. Who should be at the controls, however—humans, or artificial intelligences?

In the past few years plasma physics has begun to explore the use of modern-day data science and artificial intelligence methods to support research goals[2,3]. In this Perspective we will identify data science issues for extreme physics[4]—extremely high temperatures, densities or electromagnetic field strengths—and its unique challenges. In particular, phenomena at these conditions are highly nonlinear—small parameter changes can lead to large changes in behaviour. Interpreting extreme physics data typically requires simultaneously comprehending large amounts of complex multi-modal data from multiple different sources. Optimizing extreme physics systems requires fine-tuning over large numbers of (often highly correlated) parameters. Artificial intelligence (AI) methods have proved highly successful at teasing out correlations in large datasets like these and we believe will be crucial for understanding and optimizing systems that up to now have been inscrutable. These extreme conditions can be found in astrophysical scenarios, but can also be created using high-energy 'drivers' (often lasers) in the laboratory—millimetre-sized plasmas with temperatures and pressures higher than in the centre of the Sun. The field has seen an explosion of interest in machine learning techniques because new and future laser facilities have much higher shot (and corresponding data) rates than previous facilities. Data from laboratory experiments can help us to understand astrophysical plasmas, allow us to probe new phenomena in particle physics that conventional accelerators cannot reach, and might even lead the way to nuclear fusion as a power source. In this Review, we will highlight what data science issues are relevant for extreme plasma science, discuss successes in the field, identify what challenges remain, and look towards the future.

One of the most challenging areas of extreme plasma physics is high-energy-density physics (HEDP), a sub-field dating back to the 1940s that seeks to understand the behaviour of macroscopic matter that is simultaneously at both very high temperatures and very high pressures; typically $>10^7$ K and $>10^6$ bar. At these conditions, several

[1]Clarendon Laboratory, University of Oxford, Parks Road, Oxford, UK. [2]Lawrence Livermore National Laboratory, Livermore, CA, USA. [3]York Plasma Institute, Department of Physics, University of York, York, UK. [4]Nikhef, National Institute for Subatomic Physics, Amsterdam, The Netherlands. [5]DIFFER—Dutch Institute for Fundamental Energy Research, Eindhoven, The Netherlands. [6]Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Lisbon, Portugal. [7]Sandia National Laboratories, Albuquerque, NM, USA. [8]Imperial College London, London, UK. [9]School of Mathematics and Physics, Queen's University Belfast, Belfast, UK. [10]Laboratory for Laser Energetics, University of Rochester, Rochester, NY, USA. [11]Dutch National Center for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands. [12]AWE Plc, Aldermaston, Reading, UK. ✉e-mail: peter.hatfield@physics.ox.ac.uk; gaffney3@llnl.gov; anderson276@llnl.gov

complex areas of physics become relevant and highly coupled, making ab initio predictions very challenging. For example, in many circumstances HEDP plasmas can start to inherit properties from both 'classical plasmas' (where the behaviour of the matter can to some degree be thought of as a gas of both ions and free electrons) and condensed matter (matter at solid density where strong interactions between bound electrons are relevant)[5]. Key contemporary problems in HEDP theory include understanding multi-species plasmas, self-consistent emission, absorption and scattering of radiation, non-equilibrium plasmas, relativistic electron transport, magnetized plasmas, and quantum electrodynamic effects.

Understanding HEDP is of both great theoretical and practical importance. As already discussed, understanding these conditions is key in astrophysics. In this field, it is becoming possible to study particle and nuclear physics through HEDP experiments, as well as novel phenomena that are predicted to emerge only at extreme conditions, for example, the predicted thermal Schwinger process: the spontaneous production of electrons and positrons at very high electric field strengths[6]. Extreme physics and high-power laser science have given access to exotic forms of matter, such as new forms of ice[7] and metallic hydrogen[8]. HEDP experiments are also at the forefront of the development of new classes of particle accelerators[9] (for example, laser wakefield acceleration[10], bright γ-ray sources[11], laser-driven ion acceleration[12] and highly efficient neutron generation[13]) with wide-ranging application across science including condensed matter physics, material science and biomedical imaging[14]. Finally, the high temperatures and pressures of HEDP are one way to make nuclear fusion as a clean industrial power source a reality via inertial confinement fusion[15] (ICF).

HEDP has a rich heritage of experimentation, currently practiced by thousands of scientists in several large facilities around the world. The National Ignition Facility (NIF) at Lawrence Livermore National Laboratory (LLNL) is the most energetic laser in the world, and is the premier facility working towards ICF[16], as well as operating the exciting 'Discovery Science' programme[17,18]. At the other end of the energy spectrum are high-repetition-rate lasers (for example, Gemini at the Central Laser Facility[19]), which are much less energetic (and so can typically reach less extreme conditions), but can fire up to many times a second rather than at most a few times a day at NIF. There are many more facilities, each with unique capabilities, and a range of other technologies that are used around the world in HEDP experiments, including gas guns, Z-pinches, proton beams, pulsed power, X-ray free-electron lasers and ion accelerators (for example, at the Facility for Antiproton and Ion Research[20]). New facilities and upgrades are constantly in planning, and understanding the data we currently have is directly relevant to choices about what facilities we will need in the future. There are also important synergies with closely related areas of physics, for example, magnetic confinement fusion, solar probes and the detection of high-energy cosmic rays. Finally, alongside experimentation, huge computational facilities have traditionally been a key part of HEDP, with the development of many sophisticated simulation codes run on top high-performance computing facilities. At the time of writing, about six of the top ten supercomputers in the world are used in some capacity for simulating plasma physics experiments like ICF.

The field now needs to be able to systematically manage large quantities of data, both because the amount of experimental data is set to grow massively, and also because the capacity to simulate huge numbers of experiments is moving beyond the limits of conventional methods. The quantity of experimental data is increasing, both because shot rates on facilities are dramatically increasing (see Fig. 1), but also because diagnostics are becoming more sophisticated; around 150 GB of data are taken on each NIF shot, and Linac Coherent Light Source campaigns have reached about 70 GB per minute[21]. Machine learning[22], Bayesian methods[23] and data-driven science[24] have been much used in particle physics and astrophysics for many years, and are having large impacts on other multiscale, highly nonlinear areas of physics such as climate science and Earth system science[25,26]. Some AI solutions from other fields are likely to be applicable in plasma physics—but HEDP also has its own unique challenges. Specifically, we typically want to (sometimes very rapidly) fine tune (either optimizing or fitting a model) a large number of parameters for a desired outcome, based on a large number of multi-modal datasets. This is very difficult for humans, because it is very hard to simultaneously comprehend all the different sources and forms of data, but is achievable for AI. In HEDP, the best-performing experiments in the field are now increasingly data-driven[27,28]; the vision is to work towards a HEDP science where algorithms are at the centre of design, analysis of experiments and discovery.

In the following sections we review key challenges and topics in extreme plasma physics, and highlight research areas where data science has dramatically affected the field. We further lay out what the future might hold for data-driven 'extreme physics' and HEDP, and how the community must change and adapt its research practices to make the most of these exciting new approaches.

Challenges

The qualities that make HEDP an exciting area of research also contribute to causing the challenges inherent in any quantitative analysis of data. Table 1 summarizes some key quantitive methods, comparing conventional and emerging approaches, and in the following we discuss three primary challenges that data science techniques can help to address in HEDP.

Experimental design and automation

A key component of HEDP is experimentation. Designing experiments is a hugely complex task, and researchers will typically have a wide range of overlapping goals during this phase. Researchers must consider what specific hypotheses are to be tested, and whether the expected data would be sufficient to rule out alternatives. What design to test, diagnostic instruments to field, or astronomical observation to make will depend heavily on the specific science goal at hand. In current design approaches there is typically much use of intuition, and often experiments are built as an extension of what has been done before, limiting the regions of experimental space that are studied. Machine learning techniques offer a possible framework within which the intuition of the computational scientists and experimental scientists can be explicitly included in a cohesive picture that considers both measurements that can be made, and which aspects of physics have the most leverage on those measurements. AI-aided design is beginning to be used in the creation of new HEDP experiments[29,30], and we foresee that becoming the norm in the coming years. However, machine learning methods have yet to demonstrate that they can 'think outside the box' of pre-defined parameter spaces, so for the foreseeable future it will still be necessary to have substantial human input into the design process.

Experiments on state-of-the-art high-repetition-rate lasers, firing many times a second, cannot be done with a human in the loop, so in this case some algorithmic control is essential. This could, in principle, lead to huge savings of time, money and human effort in the near future—allowing them to be redirected to aspects of research where they can be better used. Automating experiments combines control of experimental parameters and real-time analysis of experimental results into a single algorithm. The experimental goals are coded into the automation, such that choices of how to vary the experimental inputs are made automatically in order to maximize the output.

Automation also allows for active feedback stabilization of complex processes. This is of particular benefit to HEDP experiments, where many nonlinear effects combine to determine the performance of what is a nominally unpredictable process. In this context, AI could be the solution to an otherwise intractable problem; the well known ability of deep learning models to discover complex interactions[31] could be used to optimize systems that have proved far too nonlinear for human researchers.

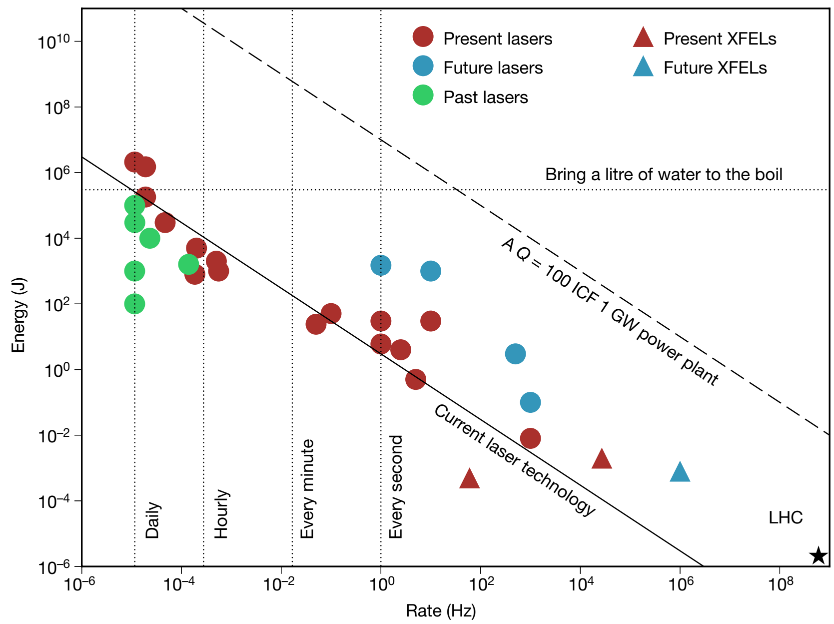

Fig. 1 | Shot rates and energy of large high-powered laser facilities in different eras. Shot rates and energies plotted are representative rather than definitive; facilities can typically operate in a number of slightly different modes, and in addition other laser properties (long pulse versus short pulse, beam colour and so on) are also important. We note also that higher shot numbers enable more parameter space to be probed, or a higher signal-to-noise ratio to be reached, but do not automatically translate into more data (depending on what diagnostics are used and the experiment performed). Facilities included are: NIF, LMJ, Omega, Gemini, Vulcan, Orion, TA2, Artemis, Vega, Titan, Texas Petawatt, HAPLS, SG-III, CoReLS 4 PW, RRCAT 150 TW, LULI2000, ALLS, Shanghai Superintense Ultrafast Laser Facility, DRACO (current), Allegra, EPAC DiPOLE, TARANIS-X, Station of Extreme Light (future), Nova, SHIVA, Cyclops, Argus, SG-II, ISKRA-V (past)[113–122], SACLA, the European XFEL (current XFELs) and LCLS-II (a future XFEL)[123–125]. The solid line illustrates the approximate state of current technology. We show for context the collision rate, and energy per collision for the LHC[126], although such figures are not directly comparable. We also indicate with a dashed line the shot rates that would be need to be achieved for an ICF power plant (for $Q = 100$; an output energy to input energy ratio of 100 per shot), to illustrate the long-term aspirations of the field.

Data synthesis

The measurements made in HEDP experiments are often highly integrated; experiments typically do not measure the actual quantity of interest, and there are usually multiple confounding or nuisance variables that must be controlled. Isolating a particular aspect of these systems is often not possible, resulting in a measurement of an evolving system subject to different conditions and physical processes. This complexity makes repeatability an issue. To isolate individual aspects of the underlying physics, experiments therefore typically require multiple, indirect observations, sometimes spread across several different experimental facilities. At most facilities researchers have developed multiple diagnostics for experiments; for example, both X-ray and particle spectra may be measured on a single experiment, along with many other forms of experimental data, all of which might contribute to the determination of a single quantity. The analysis of such increasingly sophisticated interlinked data requires the use of more advanced modelling techniques to make the best use of the available data, and to quantify the uncertainty on any inferences in a sensible manner. Looking forward, large quantities of data will require the development of streamlined automated data analysis tools to avoid read/write bottlenecks[32].

As well as combining data from multiple diagnostics, there are also challenges in combining data from multiple sources: data synthesis, where multiple forms of heterogeneous data are combined in a self-consistent manner to construct a more complete picture of the phenomenon of interest. Typically each diagnostic will be analysed separately, and overall conclusions reached heuristically by the researcher. However, combining all available data can help reduce error bars, break degeneracies and cancel out uncorrelated systematics[33].

The long-term vision for the best use of physics data is to develop systems to combine data from multiple diagnostics on the same shot, multiple shots, shots on different facilities, and finally from different types of facility.

Physics models

The evolution of HEDP experiments is governed by multiple complex, nonlinear physics models, each of which has its own range of applicability and uncertainties. Solving these multiphysics computer models requires very expensive simulations, which are often not suited to next-generation high-performance computing platforms. With the increasing desire to explore larger experimental parameter spaces, there is a shifting dynamic between single, incredibly computationally expensive, 'hero' simulations, and large-scale ensembles of simulations that become meaningful only when confronted with experimental data.

Owing to numerical approximations, poorly known or unknown model parameters, and missing physics (model discrepancy), computer models often do not accurately represent the physical process under study. We can leverage real-world experiments to calibrate our computational models, enabling us to constrain some of the uncertain model parameters; ideally including an uncertainty quantification analysis[34] and practising 'data assimilation'[35] that obeys physical laws. A useful approach to this problem is through Bayesian inversion (also known as model calibration)[36,37], which allows prior knowledge to be be included and for which convenient numerical tools exist. In a HEDP context, challenges with this approach include large parameter spaces, expensive models, and very sparse experimental data.

The computer models can often be prohibitively computationally expensive. In this case a surrogate model (or emulator) may be useful—running a moderate number of expensive simulations, and training a

Table 1 | Conventional and AI-enabled approaches to tasks in extreme plasma physics

Analytical task	Scientific task	Conventional approaches	Limitations of conventional approaches	Emergent or potential approaches
Uncertainty quantification	Quantifying uncertainty on estimates of some microphysics (such as opacity or equation of state), characterizing uncertainty estimate on laser energy required for ignition	Simulation-based sensitivity studies, basic Poisson/Gaussian uncertainties, not including correlations	Local sensitivities only, relies on estimated uncertainties in underlying parameters, cannot account for simulation bias	Markov chain Monte Carlo, dropout, bootstrapping, quantification of 'unknown unknowns', invertible neural networks
Regression	Emulating an expensive simulation, building an empirical model of microphysics	Look-up tables, polynomial regression	Struggles in high dimensions, incorporating known physics constraints, extrapolation is difficult, often not fast enough	Neural networks, Gaussian processes, autoencoders
Design	Selecting target design parameters, scheduling observation runs	Adjusting parameters by hand, using a combination of code output and designer judgement	Difficult in high dimensions, human intensive, cannot be done quickly, can miss optimal/novel designs	Bayesian optimization, genetic/evolutionary algorithms
Pattern recognition	Identifying target defects, characterizing magnetic perturbations, image segmentation/featurization	Human inspection	Laborious and time-consuming, subject to individual biases, uncontrolled approximation to true information content	Convolutional neural networks, deep learning, random forest classifiers, human–AI hybrid approaches, application-specific integrated circuits incorporated into diagnostics, generative adversarial networks
Data synthesis	Combining data from multiple sources (such as different instruments)	Researcher guided inference from independently analysed diagnostic data	Difficult to do 'by hand', difficult to take advantages of any degeneracies broken	Bayesian inference, data assimilation methods
Classification	Image classification, particle track classification, identification of good/bad shots, anomaly detection	Simple analytic criteria, human inspection	Laborious and time-consuming, potentially inaccurate	Random forest, decision trees, neural networks, deep learning, generative adversarial networks
Model calibration	Update physics models/parameters in the face of experimental data	Trial and error, single-point fitting to data, hand-tuning	Laborious and time-consuming, inaccurate or missing treatment of uncertainties, multiple solutions missed, prone to overfitting	Bayesian inference, discrepancy modelling, transfer learning, multi-task learning, physics-informed neural networks

machine learning algorithm to reproduce what the simulation would have given as an output. Surrogate models are of course themselves only an approximation of the true model, introducing further uncertainty that needs to be accounted for.

Emulation can be done at the macro level (for example, predicting outputs of a whole experiment), or at the level of individual modules run inline inside a computer model. The use of emulators opens up an array of inference methods that would not be practical with the full computational expense of a conventional simulation[38–44].

Case studies

Here we highlight three key areas on which researchers are tackling the challenges described in Section II, and where data science is strongly affecting the practice of extreme plasma science.

Astrophysics

Plasmas are found throughout the Universe: in solar physics (the centre of the Sun, the solar corona, solar wind); interplanetary, interstellar and intergalactic media; in Earth's and other planetary magnetospheres and ionospheres, and tails of comets; in compact astrophysical objects (white dwarfs, neutron stars and black holes) and their accretion disks. After direct matter–antimatter annihilation, accretion onto compact objects is the most efficient energy source in the Universe[45]; the cosmos offers plenty of opportunity to probe extreme physics.

Understanding our closest star is of course of supreme practical importance—alongside curiosity-driven blue-sky astrophysical motivation. Space weather refers to conditions on the Sun, in the heliosphere, in the solar wind, and in Earth's magnetosphere, ionosphere and thermosphere, which can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health[46,47]. Changes in the space environment, resulting mainly from changes on the Sun, include modification of the ambient plasma, particulate radiation (electrons, protons and ions), electromagnetic radiation (including radio, visible, ultraviolet, X-ray and γ radiation), and magnetic and electric fields.

As with many other areas of physics, the amount of data we have on the multi-scale complex physics experiment that is the Sun has massively increased in recent years from solar space missions for example, from both spacecraft, such as the Solar and Heliospheric Observatory (SOHO), the Hinode mission, the Solar Dynamics Observatory (SDO), the Parker Solar Probe (PSP) and Solar Orbiter (SolO), and CubeSats, such as the Miniature X-ray Solar Spectrometer (MinXSS). Machine learning has been employed successfully to both forecast and 'now-cast' space weather[48]. It has been used to make predictions and gain insight about solar wind[49], solar flares[50,51], coronal mass ejections (CMEs)[52,53], Van Allen radiation belts[54], geomagnetically induced currents[55] and the role of auroras as a proxy for ionospheric disturbances[56].

Outside our Solar System, large survey telescopes are taking huge amounts of data on astronomical bodies with extreme physics. A full range of data science based techniques are used to both identify objects of interest in the large datasets[57], and to understand their underlying physics. Machine learning and statistical methods have been used to make a Bayesian constraint on supra-nuclear equations-of-state[58], understand the interiors of exoplanets[59], constrain fundamental stellar parameters from asteroseismic observations[60], classify supernovae[61],

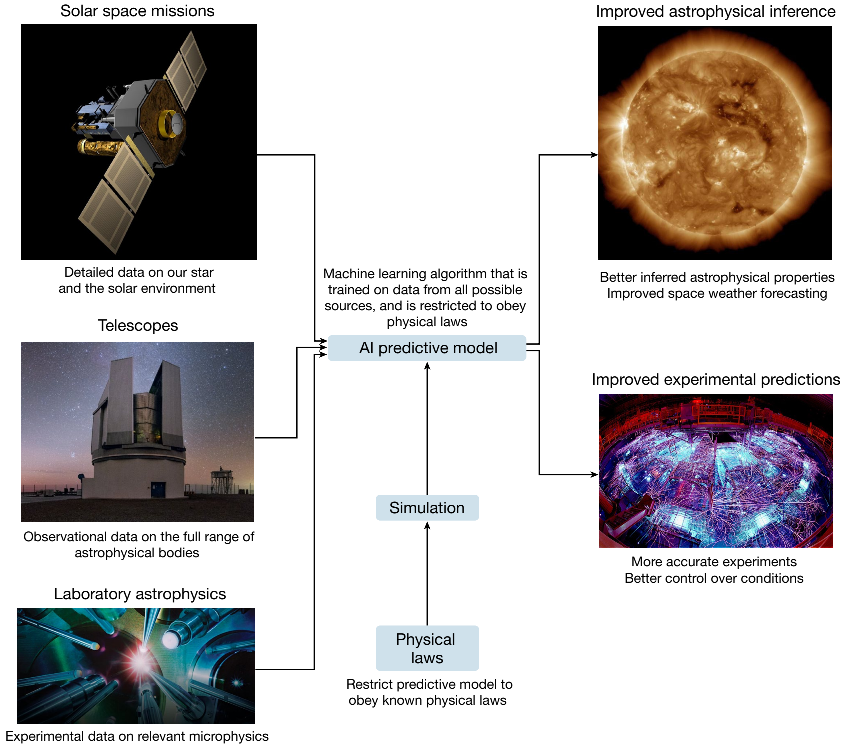

Fig. 2 | Integration of astrophysical information. Combining data from multiple sources. At the centre is a machine learning algorithm that is receiving data from multiple sources, and is able to update its beliefs based on this ('data assimilation'). The resulting data-driven results will take information both from theory and from observational and experimental data to give more precise predictions, with realistic uncertainties. Inset images courtesy of JPL/NASA/SDO/AIA (https://images.nasa.gov/details-PIA17669), JPL/ESA/NASA/SOHO (https://images.nasa.gov/details-PIA18170), ESO/Y. Beletsky (https://www.eso.org/public/images/potw2009a/) and OMEGA/LLNL (https://lasers.llnl.gov/media/photo-gallery?id=nif-1109-17880), Randy Montoya/Sandia National Laboratories (https://www.energy.gov/articles/photo-week-z-machine).

identify and infer the properties of white dwarfs[62], classify the states of black-hole X-ray binaries[63], and emulate radiative transfer during the epoch of cosmological reionization[64].

HEDP experimental facilities can, as discussed, also probe the extreme plasma physics relevant in astrophysical bodies. For example, experiments have measured the equation of state at conditions relevant for the centres of gas giant exoplanets[65,66], tested theories on possible origins of magnetic fields on galactic scales[67] and improved our understanding of white dwarf photosphere spectra[68]. There has recently been great success in applying data science methods to these experiments and making inferences with realistic uncertainties. Measurements of iron opacity at conditions relevant to solar physics on the Z facility, for example, have modelled a large number of possible sources of uncertainty and systematics, and have combined the data from multiple shots together to get realistic uncertainties. This measurement was statistically inconsistent with conventional opacity predictions and may have led to an adjustment of estimates of the metallicity (lithium and higher atomic number elements) of the Sun[69,70]. Similarly, Bayesian estimates from experiment of nuclear reaction rate at conditions relevant for Big Bang nucleosynthesis were found to be 3% different from what was conventionally assumed; a level of precision needed for cosmological studies[71]. As discussed in the introduction, future facilities will have shot rates that will make multi-shot studies like this the norm. Critically, this will make it possible to consistently make realistic uncertainty quantifications of key parameters, and will also give the ability to probe large parts of parameter space (for example, to measure the equation-of-state or opacity at a large number of points in temperature–density space). With these vast quantities of observational and experimental data, the natural next step is to use experimentally calibrated models of microphysics like these in astrophysical models—with the potential to give improved predictions over theory-only models.

There is a huge amount of data on astrophysical plasmas of a wide variety of sources, taken in a huge variety of ways, but unfortunately they are currently held in different forms, by different communities. See Fig. 2 for an infogram on how different astrophysical datasets could in future be combined in a data assimilation framework[72,73]. Astronomers have long practiced 'multi-wavelength' astronomy (astronomical observing at multiple electromagnetic wavelengths), and since about 2015 have practiced 'multi-messenger' astronomy (astronomical data-taking from multiple messengers of information, electromagnetic waves, gravitational waves, cosmic rays and neutrinos). It is now time for 'multi-provenance' astronomy; integrating data from both observations and experiments into one coherent model of our understanding of astrophysical sources.

Inertial confinement fusion

The aim of igniting a self-propagating hydrogen fusion reaction in laboratory-scale plasmas has been a scientific grand challenge for over 50 years. The motivation is clear: a reliable fusion reaction could form the basis for an effectively limitless, clean and safe energy source[74]. There are good reasons the research has been long-running; generating densities and temperatures high enough to overcome the Coulomb repulsion between nuclei and Bremsstrahlung radiation losses from electrons, over timescales long enough to allow energy break-even, is extremely difficult. In nature, the required conditions are reached in the cores of stars as a result of gravitational collapse and confinement. On Earth, we require even more extreme conditions to account for the much shorter timescales.

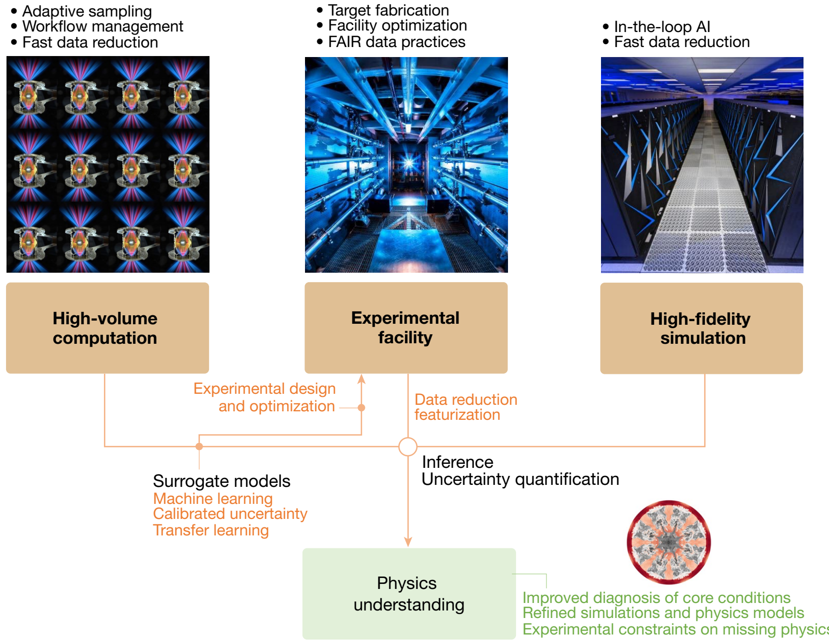

Fig. 3 | Integrating information sources in ICF studies. Our understanding of ICF implosion physics is based on a combination of high-volume, lower-fidelity simulation ensembles; sparse, difficult-to-diagnose experiments; and best-physics simulations that push the limits of high-performance computing technology. Creating and synthesizing these data into an improved understanding of the physics will require multiple complementary techniques from data science, uncertainty quantification and artificial intelligence. Inset images are courtesy of Damien Jemison/LLNL (https://www.llnl.gov/news/national-ignition-facility-makes-history-record-500-terawatt-shot), Jacob Long/LLNL (https://lasers.llnl.gov/news/rugby-hohlraum-kicks-up-nif-energy-efficiency), Randy Wong/LLNL (https://www.llnl.gov/news/lawrence-livermore-unveils-nnsa%E2%80%99s-sierra-world%E2%80%99s-third-fastest-supercomputer) and Tzanio Kolev/LLNL (https://computing.llnl.gov/projects/blast/icf-like-implosion).

In ICF studies, millimetre-scale targets are driven to implode using high-energy (tens of kJ to a few MJ) drivers[75]. For fusion ignition, the implosion and subsequent stagnation needs to generate a high temperature (10^7–10^8 K) hydrogen plasma, surrounded by a compressed (100× to 1,000×) fuel layer, confined by its own inward velocity for several hundred picoseconds[76]. Experiments on current ICF facilities like the NIF[77,78], the Omega Laser Facility[79], and the Z pulsed-power machine[80,81] routinely generate plasmas under solar core conditions and beyond.

ICF experiments present distinct data challenges owing to their scale and complexity. Experimental facilities are expensive and are not expected to achieve high repetition-rate operation any time soon. Targets and drivers are very complex resulting in high-dimensional experimental design spaces. Experiments are also highly integrated, meaning that direct measurement of any figure of merit beyond the raw energy yield is not possible. These factors mean that ICF datasets are always sparse, with multiple confounding factors and uncertain information content; researchers therefore place a very high value on theoretical studies undertaken using multiphysics simulation codes. Although cheaper than experiments, the simulations are still expensive, requiring at least months of central processing unit time to complete. They can also have substantial bias[82], and therefore require calibration against the available experimental data (from both ICF- and smaller-scale experiments focused on relevant phenomena). There is a great need for methods that can help with experimental design and optimization, interpretation of experimental data, linking experiments with physics models, as well as making reliable predictions of future experiments. As this work progresses, ICF is becoming a prototypical example of the difficulties associated with science in the data-poor regime.

As with the other examples in this Perspective, the fundamental data problem is the synthesis of multiple sources of information. Here, the key aim is to efficiently use the sparse information available to update our physics understanding in order to make simulations more predictive of future experiments. Figure 3 shows a potential workflow that fully integrates data-driven and machine-learning methods to achieve this goal. There are three fundamental information sources; experimental data which may comprise approximately 10 'shots' each producing multiple diagnostics with diverse data types; traditional high-fidelity simulations which produce single best-physics predictions for each shot; and high-volume ensemble studies which use large numbers of (necessarily lower-fidelity) simulations to investigate competing physics hypotheses and provide protection against overfitting. The optimization and control of these information sources, the combination of the resulting data, and the updating of models to become more predictive all present numerous opportunities for modern data science and machine learning approaches and each stage has seen recent active research.

Work to optimize the three information sources in Fig. 3 include the acceleration of multiphysics simulations using deep learning[41,83], infrastructure for intelligent control of large-scale simulations[84], intelligent and data-informed design of experiments[28,29,85], as well as optimization of experimental facilities[86]. These developments have enabled simulation studies of unprecedented size[87] and the generation of open-source ICF datasets[88] that motivate deep learning research[40,43,89,90]. AI tools have been applied to the automatic analysis and featurization of complex data types such as spectra, images[40,91] and line-of-sight dependent quantities. There has been much interest in using Bayesian inference to improve diagnostics[92], and to synthesize observations in both focused HEDP experiments[37] and in full-scale ICF experiments[36,93–95]. The ultimate aims of using these methods to improve physics understanding, and the reliability of simulations in extrapolating to new designs[34] or facilities, have been addressed though machine learning[28,96], Bayesian model calibration[36], and transfer learning[44,97].

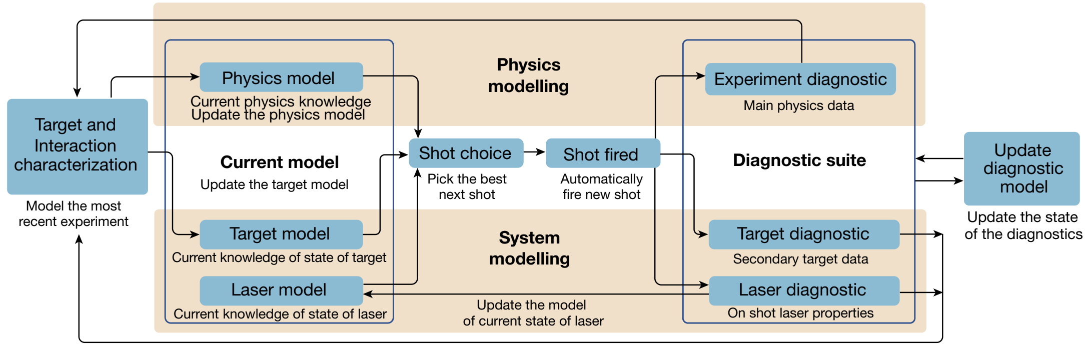

Fig. 4 | High-repetition workflow. The different components of a series of high-repetition-rate high-powered lasers are shown. The AI system must have (1) a model of its best estimate, with uncertainties, of the physics being probed, (2) a model of the current state of the laser (so that it can achieve science goals better, but also to avoid damaging itself), (3) a model of the target, (4) an algorithm to rapidly select what the next shot must be (depending on science goal), (5) a system to actually fire the shot with no human intervention, (6) rapid automated data collection, (7) rapid physics, laser and target modelling, and (8) the capacity to update a model of diagnostic performance (if needed).

Integrating the recent work we have described into a fully developed workflow similar to what is shown in Fig. 3 is still a great challenge. Once achieved, however, we expect to gain an unprecedented view of the ICF design space, much better understanding of the conditions in current ICF experiments, and an improved understanding of the path to high yield and ignition.

Automation for high repetition rates

At the other end of the spectrum from NIF, which can only fire roughly once a day, are high-repetition-rate lasers that can fire up to multiple times a second. This means taking huge amounts of data towards given science goals, but also means that large aspects of the experimental process must be automated. To succeed, the automation of experiments requires both control of experimental parameters and real-time analysis of experimental results in one single algorithmic process; see Fig. 4.

The first step in achieving this is for experimental goals to be coded in to the automation algorithm, so that choices of how to vary the experimental inputs are made automatically in order to maximize the desired output. This approach enables huge increases in efficiency in optimization and model learning experiments, which are especially important when these experiments are resource-limited. In addition, automation also allows for active feedback stabilization of complex processes; this is of particular benefit to high-energy-density and plasma physics experiments, where many nonlinear (and also often non-equilibrium) effects combine to determine the performance of what is a nominally unpredictable process. Thought must also be given to other factors that will determine optimal data return: the diagnostics used, signal-to-noise ratio achieved and the experimental parameter space covered. Many diagnostics have already been adapted for fast electronic readout[98], however, challenges still remain in adapting X-ray or charged particle imagers and similar detectors for both maximum flexibility, as well as robustness to the harsh radiation and debris environments of HEDP experiments, especially at multi-Hz repetition rates.

This approach may be used to perform the following tasks: optimization, stabilization and model inference.

(1) Optimization: a function of experimental diagnostics is used to calculate the 'fitness' that expresses how closely measurements reflect the desired performance. An iterative procedure is then performed to optimize this fitness value by controlling experiment input parameters. This can be done with any optimization procedure for example, an evolutionary approach (that is, genetic algorithms[27,99,100]), a numerical minimization method (for example, Nelder–Mead[100]) or by Bayesian optimization using a machine learned surrogate model[101,102]. This approach can also be used to limit the parameter search to satisfy some safety constraint, such as beam loss in a particle accelerator[101]. In general, applying these approaches allows for rapid optimization of experiments in a far more efficient manner than human-controlled experiments, and produces much better results.

(2) Stabilization: active feedback can improve the stability of experimental performance by rapidly controlling input parameters to counteract oscillation or drifts in the apparatus[103]. This is routinely performed to stabilize component systems, such as alignment of laser beam transport, but can also be applied to highly complex and nonlinear experimental phenomena, such as density limit disruptions in tokamaks using the predictions of neural network[104–106]. A stable output source then allows for much better experimental or source application outcomes.

(3) Model inference: a Bayesian approach to statistical inference and model validation requires incorporating experiment uncertainties from diagnostic data in a rigorous manner, accounting for correlations across all parameter spaces[36]. This can lead not only to better estimates of the uncertainty, but the results of the inference can also dramatically change. Including this approach in real time in the data-taking process can be used to ensure that the experiment optimally constrains the physical models under examination, thus getting the most information per shot (as opposed to simply the most neutrons/X-rays and so on).

Encoding the scientific goals into the algorithm making the shot choice is just one of many challenges in automating high-repetition-rate facilities. Not only the physics of interest, but also the laser system and the target setup typically have inherent nonlinearities that can make automating knowledge extraction extremely challenging. Small changes in system parameters (laser pulse width, shape, energy, focal spot conditions, target thickness and so on) can lead to large changes in experimental outcomes, requiring very fine control of the entire system. Thus the laser itself and the target must be modelled alongside the physics of interest. This complex multi-modal data must also be analysed as fast as the shot-rate to prevent another bottleneck in the experimental loop. In addition, performance of diagnostics themselves might be impaired over time (for example, if exposed to large radiation fluxes), requiring further modelling. The goal is for the AI to understand the effect of these diagnosed (and potentially undiagnosed) fluctuations in the system, rather than to be confused by it. Human intuition risks misinterpreting evidence when many parameters are changing simultaneously. Finally, data archiving from experiments will rapidly become a challenge. Substantial challenges in developing pipelines that can prevent data bottlenecks will become important, that is, the operating algorithms may have to decide whether to record or destroy data based on the quality of inputs and outputs to avoid large amounts of spurious and unimportant data occupying many terabytes of storage systems. In summary, there are two separate challenges that should not be conflated. There is both (1) the technical challenge of delivering online feedback and real-time data curation and (2) the modelling problem of automating knowledge extraction from the complex HEDP data. Researchers should take care to identify what aspects of their specific scientific problem fit into these two categories, and to seek appropriate solutions.

The computerized control of experimental parameters increases convenience for the experimental operators, allows for automatic parameter scans and reduces the likelihood of experimental errors. Enabling experiment automation requires considerable additional investment and effort in the preparation of experimental apparatus and facilities. However, time and resources spent on this endeavour yield large returns once the experiment is fully operational owing to the increase in efficiency and productivity.

Policy propositions

To help the field take advantage of the aforementioned new ways of using data, we make a series of suggestions for how educational and research practices might be beneficially changed.

Education

With the changes in HEDP data acquisition rates and analysis that have been described in this article, it is important to consider whether researchers who are new to the subject have sufficient training in the topics we have discussed.

There are courses in most universities that attempt to teach aspects of the subject; however, they are not typically included in the HEDP curriculum. At the time of writing, only three of the 40 plasma physics PhD projects, programmes and scholarships listed on http://www.findaphd.com contain a data science component.

Although the subject is easily within reach of the understanding of plasma physics graduate students, unfamiliarity with the Bayesian perspective, and the necessary jargon, render it more difficult than necessary when study is self-directed. We recommend that, as a minimum, a brief introduction[22–24] be included in one of the advanced courses, with the opportunity to take an elective course on Bayesian analysis and uncertainty quantification. Ideally, a Bayesian analysis module should be included in future HEDP doctoral programmes, built from the cases presented in this Perspective. In addition to including data science in the general curriculum, existing certificate programmes focused on both fundamental machine learning techniques and multi-disciplinary applications of data science could be leveraged by students and professionals alike. Conferences and workshops like the International Conference on Data Driven Plasma Science (which had its second meeting in 2019) and the 2018 APS Mini-Conference on Machine Learning, Data Science, and Artificial Intelligence in Plasma Research[3] have flourished in the past few years, and these opportunities should be extended to those studying or training.

For all other graduate students and researchers, a complementary approach should be taken, through targeted workshops and schools. We believe that because the field is developing very quickly, it is preferable to consider teaching these topics as part of the existing HEDP workshops that have been established over the past few years for those new to the area. Topics should include Bayesian and frequentist statistics comparisons, uncertainty quantification, Markov chain Monte Carlo sampling, surrogate building, deep neural nets, and optimization techniques for a range of dimensional spaces. Such workshops and meetings may also provide opportunities to engage the general machine learning community. The HEDP community could also entice machine learning researchers to collaborate in our field by presenting or organizing focused sessions at machine learning conferences, releasing datasets, exploring a machine learning challenge call focused on an application in our field, and pursuing direct research collaborations. Finally, we note that data science skill sets have broad application both in other research areas, as well as in industry. The proposed training is highly transferable and promises to be valuable to students regardless of specific career goals.

Research practices

The changing nature of the field means that practices within the field will also have to change. Researchers will need to become familiar with methods needed to run large numbers of simulations, and tools for storing larger amounts of data[32]. In particular, the field will have to develop data standards so that data are easily compatible between different facilities. Adopting open data practices[88] (such as FAIR: findability, accessibility, interoperability and reusability[107]) wherever possible is also likely to greatly aid collaboration and comparison of datasets, although this will not be possible for all researchers in this area. The importance of these approaches is already well understood by researchers in other fields, such as high-energy physics[108] and astronomy[109]; we foresee that a similar level of data curation will soon be necessary in HEDP.

Many of the methods described in this paper require large computational resources and good synthetic diagnostics, that is, good simulations of what the data should look like through the actual pipeline the real data goes through. While most research efforts include a computational component, these are often disjoint from the analysis of experimental data. In future it is advisable for experimental time to also have associated funding for computation and analysis, and the development of synthetic diagnostics/data analysis. Diagnostics and experiments should be designed so that collected data can easily be used in conjunction with other shots (for example, consistent pixel sizes). It may even be necessary that the commissioning of instruments comes with a corresponding budget to develop tools to simulate data as seen by the device. Data analysis is a core component of project commissioning and planning in many other areas of physics (for example, the Euclid telescope[110]).

High data rates are permitting the probing of low signal-to-noise phenomena, for example, possible physics beyond the standard model, like axions. If laser-based accelerators are to play a part in the future of probing new physics in this way, then statistical analysis must be brought up to the same standards as are practised in high-energy physics. In particular, we may wish to adopt stringent statistical significance requirements, for example, 5σ for any discovery of physics beyond the standard model[111]. Similarly, the use of blinding methods is likely to become more necessary, where researchers deliberately hide some aspect of the labelling of the data from themselves to prevent subconscious bias or p-hacking[112]. Using a blinding protocol also presents dangers and challenges; the complexity and bespoke properties common in many laser-plasma experiments make blinding difficult to implement. However, high-repetition rates will make blinding strategies much more viable, and may help in analyses where particular outcomes have great psychological importance or there are lots of different potentially viable approaches to doing the analysis.

Conclusion

The volume of data on plasma physics at the extremes is rapidly growing and offers the potential for a dramatically increased speed of scientific advance. These data are typically multi-modal and describe very nonlinear systems, making interpretation challenging for humans, but tractable for AI algorithms. Plasma physics is unlikely to reach the extremely high data rates of high-energy physics in the immediate future, and AI will not be able to solve all problems in the field. Nonetheless, data science offers novel ways of working and ways to gain insight; we hope practitioners in the field will be able to find applicability for these methods in their research: see Box 1.

In conclusion, modern data science has a lot to offer extreme plasma physics and HEDP science; the community must act now to identify

Box 1

Key conclusions and recommendations

Conclusions

(1) The application of machine learning and modern data science methods to extreme plasma physics and HEDP is rapidly growing and is helping to produce realistic uncertainties on predictions.
(2) Higher-repetition-rate facilities have opened up a range of ways of working; data-driven discovery, blinding methods, greater reproducibility and automated data taking.
(3) Integrating machine-learning-based approaches into working practices can save much money, time and human effort.
(4) AI-based tools are now often more successful at optimizing nonlinear extreme physics systems and comprehending multi-modal data than humans.

Recommendations

(1) Researchers should think carefully about how best to use their data: what methods and diagnostics they can use to take the best data, get sensible uncertainties, and coherently combine with other datasets.
(2) Awards of experimental time and instrument construction should also include greater support for uncertainty quantification, building synthetic diagnostics and data analysis.
(3) Plasma physics graduate education and national laboratory training programmes should begin to include basic data science courses.
(4) Researchers should try where possible to practice open science best practice; making code and data available publicly, using shared data standards between different facilities.

which areas it will make the biggest impact in and make resources and training available to make the most of these novel approaches.

1. Rose, S. Set the controls for the heart of the Sun. *Contemp. Phys*. **45**, 109–121 (2004).
2. Spears, B. K. et al. Deep learning: a guide for practitioners in the physical sciences. *Phys. Plasmas* **25**, 080901 (2018).
This tutorial paper gives an introduction to scientific machine learning, with examples taken from ICF research.
3. Wang, Z., Peterson, J. L., Rea, C. & Humphreys, D. Special issue on machine learning, data science, and artificial intelligence in plasma research. *IEEE Trans. Plasma Sci*. **48**, 1–2 (2020).
4. Colvin, J. & Larsen, J. *Extreme Physics* (Cambridge Univ. Press, 2013).
5. Graziani, F., Desjarlais, M. P., Redmer, R. & Trickey, S. B. (eds) *Frontiers and Challenges in Warm Dense Matter* Lecture Notes in Computational Science and Engineering Vol. 96 (Springer, 2014).
6. Gould, O., Mangles, S., Rajantie, A., Rose, S. & Xie, C. Observing thermal Schwinger pair production. *Phys. Rev. A* **99**, 052120 (2019).
7. Millot, M. et al. Experimental evidence for superionic water ice using shock compression. *Nat. Phys*. **14**, 297–302 (2018).
8. Celliers, P. M. et al. Insulator–metal transition in dense fluid deuterium. *Science* **361**, 677–682 (2018).
9. Joshi, C. & Malka, V. Focus on laser- and beam-driven plasma accelerators. *New J. Phys*. **12**, 045003 (2010).
10. Hidding, B., Foster, B., Hogan, M. J., Muggli, P. & Rosenzweig, J. B. Directions in plasma wakefield acceleration. *Phil. Trans. R. Soc. Math. Phys. Eng. Sci*. **377**, 20190215 (2019).
11. Wang, W.-M. et al. Collimated ultrabright gamma rays from electron wiggling along a petawatt laser-irradiated wire in the QED regime. *Proc. Natl Acad. Sci. USA* **115**, 9911–9916 (2018).
12. Badziak, J. Laser-driven ion acceleration: methods, challenges and prospects. *J. Phys. Conf. Ser*. **959**, 012001 (2018).
13. Feng, J. et al. High-efficiency neutron source generation from photonuclear reactions driven by laser plasma accelerator. *High Energy Density Phys*. **36**, 100753 (2020).
14. Hidding, B. et al. Plasma wakefield accelerator research 2019–2040: a community-driven UK roadmap compiled by the plasma wakefield accelerator steering committee (PWASC). Working Paper (2019); preprint at https://arxiv.org/abs/1904.09205.
15. Nuckolls, J., Wood, L., Thiessen, A. & Zimmerman, G. Laser compression of matter to super-high densities: thermonuclear applications. *Nature* **239**, 139–142 (1972).
16. Hurricane, O. A. & Program, I. Overview of progress and future prospects in indirect drive implosions on the National Ignition Facility. *J. Phys. Conf. Ser*. **717**, 012005 (2016).
17. Fournier, K. *NIF Monthly Highlights for July 2019*. LLNL Technical Report, LLNL-TR-785259; NIF-1006466629981592, http://www.osti.gov/servlets/purl/1548376/ (Lawrence Livermore National Laboratory, 2019).
18. Remington, B. A., Rudd, R. E. & Wark, J. S. From microjoules to megajoules and kilobars to gigabars: probing matter at extreme states of deformation. *Phys. Plasmas* **22**, 090501 (2015).
19. Trines, R. *CLF Annual Report 2017–2018*, https://www.clf.stfc.ac.uk/Pages/Annual-Report-2017-18.aspx (Science and Technology Facilities Council, 2019).
20. Sturm, C. & Stöcker, H. The facility for antiproton and ion research FAIR. *Phys. Part. Nucl. Lett*. **8**, 865–868 (2011).
21. MacDonald, M. J. et al. Measurement of high-dynamic range X-ray Thomson scattering spectra for the characterization of nano-plasmas at LCLS. *Rev. Sci. Instrum*. **87**, 11E709 (2016).
22. Mitchell, T. *Machine Learning* (McGraw-Hill, 1997).
23. Sivia, D. & Skilling, J. *Data Analysis: A Bayesian Tutorial* (Oxford Univ. Press).
24. Brunton, S. L. & Kutz, J. N. *Data-Driven Science and Engineering* (Cambridge Univ. Press, 2019).
25. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
26. Fleming, S. W. & Gupta, H. V. The physics of river prediction. *Phys. Today* **73**, 46–52 (2020).
27. Streeter, M. J. V. et al. Temporal feedback control of high-intensity laser pulses to optimize ultrafast heating of atomic clusters. *Appl. Phys. Lett*. **112**, 244101 (2018).
28. Gopalaswamy, V. et al. Tripled yield in direct-drive laser fusion through statistical modelling. *Nature* **565**, 581–586 (2019).
This paper is the first to use data-driven approaches to motivate and carry out new ICF experiments.
29. Hatfield, P. W., Rose, S. J. & Scott, R. H. H. The blind implosion-maker: automated inertial confinement fusion experiment design. *Phys. Plasmas* **26**, 062706 (2019).
30. Martin, M., London, R., Goluoglu, S. & Whitley, H. An automated design process for short pulse laser driven opacity experiments. *High Energy Density Phys*. **26**, 26–37 (2018).
31. Raghu, M. & Schmidt, E. A Survey of Deep Learning for Scientific Discovery (2020); preprint at http://arxiv.org/abs/2003.11755.
32. Thayer, J. et al. Data systems for the Linac Coherent Light Source. *Adv. Struct. Chem. Imag*. **3**, 3 (2017).
33. Bernal, J. L., Peacock, J. A., Bernal, J. L. & Peacock, J. A. Conservative cosmology: combining data with allowance for unknown systematics. *J. Cosmol. Astroparticle Phys*. **2018**, 002 (2018).
This paper considers how to get realistic uncertainty estimates in the presence of unknown systematics.
34. Osthus, D., Vander Wiel, S. A., Hoffman, N. M. & Wysocki, F. J. Prediction uncertainties beyond the range of experience: a case study in inertial confinement fusion implosion experiments. *SIAM/ASA J. Uncertainty Quant*. **7**, 604–633 (2019).
35. Lang, M. & Owens, M. J. A variational approach to data assimilation in the solar wind. *Space Weather* **17**, 59–83 (2019).
36. Gaffney, J. A. et al. Making inertial confinement fusion models more predictive. *Phys. Plasmas* **26**, 082704 (2019).
37. Kasim, M. F., Galligan, T. P., Topp-Mugglestone, J., Gregori, G. & Vinko, S. M. Inverse problem instabilities in large-scale modeling of matter in extreme conditions. *Phys. Plasmas* **26**, 112706 (2019).
38. van de Plassche, K. L. et al. Fast modeling of turbulent transport in fusion plasmas using neural networks. *Phys. Plasmas* **27**, 022310 (2020).
39. Meneghini, O. et al. Self-consistent core-pedestal transport simulations with neural network accelerated models. *Nucl. Fusion* **57**, 086034 (2017).
40. Anirudh, R., Thiagarajan, J. J., Bremer, P.-T. & Spears, B. K. Improved surrogates in inertial confinement fusion with manifold and cycle consistencies. *Proc. Natl Acad. Sci. USA* **117**, 9741–9746 (2020).
41. Kluth, G. et al. Deep learning for NLTE spectral opacities. *Phys. Plasmas* **27**, 052707 (2020).
This paper uses machine learning to emulate opacity calculations, in order to speed up ICF simulations.
42. Humbird, K. D., Peterson, J. L. & McClarren, R. G. Deep neural network initialization with decision trees. *IEEE Trans. Neural Netw. Learn. Syst*. **30**, 1286–1295 (2019).
This paper presents a data-driven method for selecting deep neural network architecture and hyperparameters
43. Kasim, M. F. et al. Building high accuracy emulators for scientific simulations with deep neural architecture search. Preprint at https://arxiv.org/abs/2001.08055 (2020).
44. Kustowski, B. et al. Transfer learning as a tool for reducing simulation bias: application to inertial confinement fusion. *IEEE Trans. Plasma Sci*. **48**, 46–53 (2020).
Transfer learning updates a subset of a pre-trained deep neural network using experimental data; this paper explores its application to sparse ICF datasets.
45. Bambi, C. Astrophysical black holes: a compact pedagogical review. *Ann. Phys*. **530**, 1700430 (2018).
46. *The National Space Weather Program Strategic Plan*. Technical Report CM-P30–1995, https://www.icams-portal.gov/publications/spacewx/nswp2.html (ICAMS, 1995).
47. Eastwood, J. P. et al. The economic impact of space weather: where do we stand? *Risk Anal*. **37**, 206–218 (2017).
48. Camporeale, E. The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weather* **17**, 1166–1207 (2019).
49. Camporeale, E., Carè, A. & Borovsky, J. E. Classification of solar wind with machine learning. *J. Geophys. Res. Space Phys*. **122**, 10910–10920 (2017).
50. Chen, Y. et al. Identifying solar flare precursors using time series of SDO/HMI images and SHARP parameters. *Space Weather* **17**, 1404–1426 (2019).
51. Campi, C. et al. Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence. *Astrophys. J*. **883**, 150 (2019).

52. Inceoglu, F. et al. Using machine learning methods to forecast if solar flares will be associated with CMEs and SEPs. *Astrophys. J.* **861**, 128 (2018).
53. Bobra, M. G. & Ilonidis, S. Predicting coronal mass ejections using machine learning methods. *Astrophys. J.* **821**, 127 (2016).
54. Sarma, R. et al. Bayesian inference of quasi-linear radial diffusion parameters using Van Allen probes. *J. Geophys. Res. Space Phys.* **125**, e2019JA027618 (2020).
55. Camporeale, E. et al. A gray-box model for a probabilistic estimate of regional ground magnetic perturbations: enhancing the NOAA operational geospace model with machine learning. *J. Geophys. Res. Space Phys.* **125**, e27684 (2020).
56. Lamb, K. et al. Correlation of auroral dynamics and GNSS scintillation with an autoencoder. In *Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019)* (2019); preprint at https://arxiv.org/abs/1910.03085.
57. Rowlinson, A. et al. Identifying transient and variable sources in radio images. *Astron. Comput.* **27**, 111–129 (2019).
58. Coughlin, M. W., Dietrich, T., Margalit, B. & Metzger, B. D. Multi-messenger Bayesian parameter inference of a binary neutron-star merger. *Mon. Not. R. Astron. Soc. Lett.* **489**, L91–L96 (2019).
This paper uses both gravitational-wave and multi-wavelength electromagnetic data to simultaneously constrain stellar masses, orbital parameters, and supra-nuclear density equations of state.
59. Dorn, C. et al. A generalized Bayesian inference method for constraining the interiors of super Earths and sub-Neptunes. *Astron. Astrophys.* **597**, A37 (2017).
60. Bellinger, E. P. et al. Fundamental parameters of main-sequence stars in an instant with machine learning. *Astrophys. J.* **830**, 31 (2016).
61. Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O. & Winter, M. K. Photometric supernova classification with machine learning. *Astrophys. J. Suppl.* **225**, 31 (2016).
62. Kong, X. et al. Spectral feature extraction for DB white dwarfs through machine learning applied to new discoveries in the SDSS DR12 and DR14. *Publ. Astron. Soc. Pacif.* **130**, 084203 (2018).
63. Huppenkothen, D., Heil, L. M., Hogg, D. W. & Mueller, A. Using machine learning to explore the long-term evolution of GRS 1915+105. *Mon. Not. R. Astron. Soc.* **466**, 2364–2377 (2017).
64. Chardin, J. et al. A deep learning model to emulate simulations of cosmic reionization. *Mon. Not. R. Astron. Soc.* **490**, 1055–1065 (2019).
65. Saumon, D. & Guillot, T. Shock compression of deuterium and the interiors of Jupiter and Saturn. *Astrophys. J.* **609**, 1170–1180 (2004).
66. Smith, R. F. et al. Ramp compression of diamond to five terapascals. *Nature* **511**, 330–333 (2014).
67. Tzeferacos, P. et al. Laboratory evidence of dynamo amplification of magnetic fields in a turbulent plasma. *Nat. Commun.* **9**, 591 (2018).
68. Falcon, R. E. et al. Laboratory measurements of white dwarf photospheric spectral lines: Hβ. *Astrophys. J.* **806**, 214 (2015).
69. Bailey, J. E. et al. A higher-than-predicted measurement of iron opacity at solar interior temperatures. *Nature* **517**, 56–59 (2015).
70. Nagayama, T. et al. Systematic study of L-shell opacity at stellar interior temperatures. *Phys. Rev. Lett.* **122**, 235001 (2019).
71. de Souza, R. S., Boston, S. R., Coc, A. & Iliadis, C. Thermonuclear fusion rates for tritium + deuterium using Bayesian methods. *Phys. Rev. C* **99**, 014619 (2019).
72. Zhelavskaya, I. S., Shprits, Y. Y. & Spasojević, M. Empirical modeling of the plasmasphere dynamics using neural networks. *J. Geophys. Res. Space Phys.* **122**, 11227–11244 (2017).
73. Zhelavskaya, I. S., Shprits, Y. Y. & Spasojevic, M. in *Machine Learning Techniques for Space Weather* 301–327 (Elsevier, 2018).
74. Freidberg, J. *Plasma Physics and Fusion Energy* (Cambridge Univ. Press, 2008).
75. Atzeni, S., Meyer-ter Vehn, J. & Meyer-ter Vehn, J. *The Physics of Inertial Fusion: BeamPlasma Interaction, Hydrodynamics, Hot Dense Matter* International Series of Monographs on Physics (Clarendon Press, 2004).
76. Lindl, J. *Inertial Confinement Fusion: The Quest for Ignition and Energy Gain Using Indirect Drive* (AIP Press, 1998).
77. Campbell, E. M. & Hogan, W. J. The National Ignition Facility—applications for inertial fusion energy and high-energy-density science. *Plasma Phys. Contr. Fusion* **41**, B39–B56 (1999).
78. Moses, E. I. Ignition on the National Ignition Facility. *J. Phys. Conf. Ser.* **112**, 012003 (2008).
79. Boehly, T. et al. Initial performance results of the OMEGA laser system. *Opt. Commun.* **133**, 495–506 (1997).
80. Slutz, S. A. High-gain magnetized inertial fusion. *Phys. Rev. Lett.* **108**, 025003 (2012).
81. Gomez, M. R. et al. Experimental demonstration of fusion-relevant conditions in magnetized liner inertial fusion. *Phys. Rev. Lett.* **113**, 155003 (2014).
82. Dimonte, G. Quantitative metrics for evaluating thermonuclear design codes and physics models applied to the National Ignition Campaign. *Phys. Plasmas* **27**, 052709 (2020).
83. Yang, C. et al. *Preparing Dense Net for Automated HYDRA Mesh Management via Reinforcement Learning*. Technical Report LLNL-TR-799958, https://www.osti.gov/servlets/purl/1580017 (Lawrence Livermore National Laboratory, 2019).
This report details the use of deep neural networks to automatically perform adaptive mesh refinement in radiation-hydrodynamics simulations.
84. Peterson, J. L. et al. Merlin: enabling machine learning-ready HPC ensembles. Preprint at https://arxiv.org/abs/1912.02892 (2019).
85. Peterson, J. L. et al. Zonal flow generation in inertial confinement fusion implosions. *Phys. Plasmas* **24**, 032702 (2017).
This paper presents the first novel ICF design reached using machine learning.
86. Amorin, C., Kegelmeyer, L. M. & Kegelmeyer, W. P. A hybrid deep learning architecture for classification of microscopic damage on National Ignition Facility laser optics. *Stat. Analysis Data Mining* **12**, 505–513 (2019).
Here, deep neural network-based image classification is used to detect and classify damage in the NIF laser system, allowing more reliable operation at high energy.
87. Nora, R., Peterson, J. L., Spears, B. K., Field, J. E. & Brandon, S. Ensemble simulations of inertial confinement fusion implosions. *Stat. Analysis Data Mining* **10**, 230–237 (2017).
88. Gaffney, J. A. et al. *The JAG Inertial Confinement Fusion Simulation Dataset For Multi-Modal Scientific Deep Learning* https://library.ucsd.edu/dc/object/bb5534097t (2020).
This dataset is an example of open data practices in ICF, and is one of the first multi-modal scientific datasets to be released by the HEDP community.
89. Thiagarajan, J. J. et al. Designing accurate emulators for scientific processes using calibration-driven deep models. *Nat. Commun.* **11**, 5622 (2020).
90. Hatfield, P. et al. Using sparse Gaussian processes for predicting robust inertial confinement fusion implosion yields. *IEEE Trans. Plasma Sci.* **48**, 14–21 (2020).
91. Glinsky, M. E. et al. Quantification of MagLIF morphology using the Mallat scattering transformation. *Phys. Plasmas* **27**, 112703 (2019).
This paper uses a novel, physics-motivated deep neural network architecture to featurize images of ICF implosions driven on the Z pulsed-power machine.
92. Palaniyappan, S. et al. Hydro-scaling of direct-drive cylindrical implosions at the OMEGA and the National Ignition Facility. *Phys. Plasmas* **27**, 042708 (2020).
93. Gaffney, J., Clark, D., Sonnad, V. & Libby, S. Development of a Bayesian method for the analysis of inertial confinement fusion experiments on the NIF. *Nucl. Fusion* **53**, 073032 (2013).
94. Gaffney, J., Clark, D., Sonnad, V. & Libby, S. Bayesian inference of inaccuracies in radiation transport physics from inertial confinement fusion experiments. *High Energy Density Phys.* **9**, 457–461 (2013).
95. Knapp, P. *A Bayesian Parameter Estimation Framework for Understanding Fusion Experiments on Z*. Technical Report SAND2018-1698PE663774, https://www.osti.gov/biblio/1525597 (2018).
96. Hsu, A., Cheng, B. & Bradley, P. A. Analysis of NIF scaling using physics informed machine learning. *Phys. Plasmas* **27**, 012703 (2020).
97. Humbird, K. D., Peterson, J. L., Spears, B. K. & McClarren, R. G. Transfer learning to model inertial confinement fusion experiments. *IEEE Trans. Plasma Sci.* **48**, 61–70 (2020).
98. Götzfried, J. et al. Research towards high-repetition rate laser-driven X-ray sources for imaging applications. *Nucl. Instrum. Meth. Phys. Res. A* **909**, 286–289 (2018).
99. He, Z.-H. et al. Coherent control of plasma dynamics. *Nat. Commun.* **6**, 7156 (2015).
100. Dann, S. J. et al. Laser wakefield acceleration with active feedback at 5 Hz. *Phys. Rev. Accel. Beams* **22**, 041303 (2019).
101. Kirschner, J., Mutny, M. M., Hiller, N., Ischebeck, R. & Krause, A. Adaptive and safe Bayesian optimization in high dimensions via one-dimensional subspaces. In *Proc. Machine Learning Res.* **97**, 3429–3438 (2019).
This paper describes a Bayesian optimization method for a free electron laser that prevents the laser from violating any safety constraints.
102. Shalloo, R. J. et al. Automation and control of laser wakefield accelerators using Bayesian optimization. *Nat. Commun.* **11**, 6355 (2020).
103. Maier, A. R. et al. Decoding sources of energy variability in a laser-plasma accelerator. *Phys. Rev. X* **10**, 031039 (2020).
104. Team, J. T. et al. Hybrid neural network for density limit disruption prediction and avoidance on J-TEXT tokamak. *Nucl. Fusion* **58**, 056016 (2018).
105. Fu, Y. et al. Machine learning control for disruption and tearing mode avoidance. *Phys. Plasmas* **27**, 022501 (2020).
106. Kates-Harbeck, J., Svyatkovskiy, A. & Tang, W. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568**, 526–531 (2019).
107. Wilkinson, M. D. et al. Comment: The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
108. Albertsson, K. et al. Machine learning in high energy physics community white paper. *J. Phys. Conf. Ser.* **1085**, 022008 (2018).
109. Borne, K. D. et al. The revolution in astronomy education: data science for the masses. Preprint at https://arxiv.org/abs/0909.3895 (2009).
110. Pasian, F. et al. Science ground segment for the ESA Euclid Mission. In *Software and Cyberinfrastructure for Astronomy II* (eds Radziwill, N. M. & Chiozzi, G.) Vol. 8451, 845104 (SPIE, 2012).
111. Lyons, L. Discovering the significance of 5 sigma. Preprint at https://arxiv.org/abs/1310.1284 (2013).
This paper discusses the background, rationale, and advantages and disadvantages of the 5σ criterion commonly used in particle physics.
112. Roodman, A. Blind analysis in particle physics. In *Proc. Conf. on Statistical Problems in Particle Physics, Astrophysics and Cosmology* (SLAC, 2003).
113. Roso, L. High repetition rate petawatt lasers. *EPJ Web Conf.* **167**, 01001 (2018).
114. Zheng, W. et al. Laser performance of the SG-III laser facility. In *High Power Laser Science and Engineering* Vol. 4, e21 (Cambridge Univ. Press, 2016).
115. Danson, C. N. et al. Petawatt and exawatt class lasers worldwide. In *High Power Laser Science and Engineering* Vol. 7, e54 (Cambridge Univ. Press, 2019).
This paper reviews some of the many high-powered lasers in use around the world.
116. *Opportunities in Intense Ultrafast Lasers* https://doi.org/10.17226/24939 (National Academies Press, 2018).
117. Lin, Z. et al. SG-II laser elementary research and precision SG-II program. *Fusion Eng. Des.* **44**, 61–66 (1999).
118. *LULI2000 User Guide*. https://gargantua.polytechnique.fr/siatel-web/app/linkto/mICYYYTJle5S (Laboratoire pour l'Utilisation des Lasers Intenses, 2019).
119. Kirillov, G. A., Murugov, V. M., Punin, V. T. & Shemyakin, V. I. High power laser system ISKRA V. *Laser Part. Beams* **8**, 827–831 (1990).
120. Zhao, Z. & Wang, D. XFEL Projects in China. In *Proc. LINAC2018 (Beijing)* http://accelconf.web.cern.ch/linac2018/html/author.htm (2019).
121. Zhang, Z. et al. The laser beamline in SULF facility. In *High Power Laser Science and Engineering* Vol. 8, e4 (Cambridge Univ. Press, 2020).
122. Schramm, U. et al. First results with the novel petawatt laser acceleration facility in Dresden. In *J. Phys. Conf. Ser.* **874**, 12028 (2017).
123. European XFEL. *Facts and Figures* https://www.xfel.eu/facility/overview/facts_amp_figures/index_eng.html.
124. Linac Coherent Light Source. *LCLS-II Design & Performance* https://lcls.slac.stanford.edu/lcls-ii/design-and-performance.

125. Yabashi, M., Tanaka, H., Tono, K. & Ishikawa, T. Status of the SACLA facility. *Appl. Sci*. **7**, 604 (2017).
126. CERN. *FAQ – LHC the Guide*. Technical Report CERN-Brochure-2017–002-Eng, https://cds.cern.ch/record/2255762 (CERN, 2017).

Acknowledgements This Perspective is the result of a meeting at the Lorentz Center, University of Leiden, 13–17 January 2020. The Lorentz Centre is funded by the Dutch Research Council (NWO) and the University of Leiden. The meeting also had support from the John Fell Oxford University Press (OUP) Research Fund. The organizers are grateful to T. Uitbeijerse (Lorentz Center) for facilitating the meeting. P.W.H. acknowledges funding from the Engineering and Physical Sciences Research Council. A portion of this work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. J.A.G. and G.J.A. were supported by LLNL Laboratory Directed Research and Development project 18-SI-002. The paper has LLNL tracking number LLNL-JRNL-811857. This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the US Department of Energy or the United States Government.

Author contributions P.W.H., J.A.G. and G.J.A. conceived the work and led the writing of the manuscript. All authors contributed to the manuscript and the ideas discussed at the Lorentz Center Meeting.

Competing interests The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to P.W.H., J.A.G. or G.J.A.
Peer review information *Nature* thanks Paul Bradley, Michael Bussmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.