Adoption of Ethical Principles for Machine Learning in the Meteorological Domain

DE_398_EVIDEN-DE-Ethical-Machine-Learning

Issued by: Atos / Author names: Djordje Benn-Maksimovic, Sebastian Schmidt, Ranjith Mahendran

Date: 29.01.2026

Ref: DE_398_D398.7.5.1_202512_Adoption_Ethical_ML_Principles_v1.0.docx

Official reference number service contract: 2024/DE_398_EVIDEN

Status: Public

Funded by the European Union Destination Earth implemented by ECMWF esa EUMETSAT

Change Log

Version	Date	Description
0.1	03.11.2025	First draft of Introduction and Conclusion
0.2	01.12.2025	First draft
0.9	18.12.2025	Draft for final review
1.0	29.01.2025	Final version

Reviewers

Name	Organisation	Expertise
Stephan Siemen	ECMWF	Strategic Project Coordinator - Leading various strategic activities across ECMWF, e.g. AI Ethics
Jan Ruhnke	Atos	Leading research projects for public sector clients, e.g. for robustness of AI systems against adversarial attacks Previously Director AI of the Artificial Intelligence Centre Hamburg (ARIC e.V.)
Andrei Nutas	Atos	AI Quality Assurance Engineer with research background in AI ethics and the philosophy of technology

Contents

Glossary

AI	Artificial Intelligence
AIFS	ECMWF’s Artificial Intelligence/Integrated Forecasting System
DestinE	Destination Earth initiative
DEDL	DestinE Data Lake
DESP	DestinE Core Service Platform
DT	DestinE Digital Twins
ECMWF	European Centre for Medium-Range Weather Forecasts
EU	European Union
EUMETNET	European Meteorological Network[1]
GDPR	General Data Protection Regulation
HPC	High-Performance Computing
ICON	NWP model by German Weather Service (DWD)
IFS	ECMWF’s Integrated Forecasting System
LLM	Large Language Model
ML	Machine Learning
NWP	Numerical Weather Prediction
RMSE	Root Mean Square Error
WMO	World Meteorological Organization

[1] https://www.eumetnet.eu

1 Executive Summary

Weather forecasts and climate prognoses play a vital role in many day-to-day decision-making processes by public authorities and industry alike. They are crucial to allow our society to adapt to extreme weather events and the impact of the changing climate. The utility of these forecasts depends strongly on decision makers' trust in these predictions, which is not just related to quality of forecasts but also the alignment with values and regulations within the process making decisions. The rapid advances of data-driven weather forecasting models and the beginning of a wider spread of their use, does not only come with questions about their capabilities but also about ethical considerations of leveraging these newer technologies compared with more traditional physics-based approaches.

The European Union's Destination Earth flagship initiative (DestinE) is a key driver in demonstrating and deploying new AI-based capabilities to support adaptation measures for extreme weather events and climate change. This whitepaper presents work carried out under DestinE to explore the ethical challenges and propose avenues and guidance. This will support AI related developments by ensuring they are based on solid ethical foundations and reduce any delays in putting these developments in impactful decision-making processes.

The aim is to develop best practices to maintain the ethical development and usage of these models and ensure ongoing trust in weather forecasts, including possibilities to operationalize ethical aspects during development. The focus will be on new ethical aspects of data-driven forecasting compared to NWP-based models and discuss how consideration of ethical issues relates to trust of users and providers of various downstream applications. The DestinE initiative serves as context to illustrate these issues and how to address them.

Key takeaways include:

- AI is improving weather and climate forecasts and changes the way users can interact with these, helping many impact sectors like energy, agriculture, aviation, and disaster response.
- Trustworthy AI systems are crucial for public trust and long-term adoption, where the core dimensions of trustworthiness for meteorological applications are reliability, safety & security, fairness, and transparency.
- Supporting users of AI systems to build competences and confidence to use and apply AI in their decision-making processes and a community of best practices are essential, and the work of this study and its white papers will serve as guidance.
- DestinE and, in general, decision making around weather and climate predictions are important use cases that can be used to test the developing ethical landscape for AI and provide feedback for future guidance.

The whitepaper serves as the opener for a series of whitepapers providing an overview of challenges and approaches and exploring these topics more deeply.

2 Introduction

The integration of Artificial Intelligence (AI) into weather and climate forecasting marks a significant evolution from traditional modelling approaches. Conventional Numerical Weather Prediction (NWP), which relies on simulations explicitly modelling physical laws, is increasingly being supplemented by data-driven solutions based on Machine Learning (ML). Due to potentially lower computational costs, this shift promises substantial improvements in forecast accuracy, timeliness, and granularity that were previously unattainable [1]. These advancements can lead to better decision-making in critical sectors such as transport, energy and urban planning, for better disaster preparedness and climate change adaptation.

Complementing these innovations, the European Union's Destination Earth (DestinE) initiative aims to create high-precision digital models of the Earth to monitor and predict environmental changes at unprecedented spatial resolution and decadal timescales. The frameworks developed within the initiative open the door for a multitude of downstream applications from key impact sectors like renewable energy, public health, fishery and agriculture where ML-based algorithms can be deployed. The initiative's initial focus on two high-priority digital twins for scenario simulation for climate change adaptation and the prediction of weather-induced and geophysical extremes exemplify the high stakes and potential involved in these systems.

On the other hand, data-driven methods come with their own set of challenges. While trust has always been indispensable in weather forecasting, the introduction of AI into this field amplifies these needs due to the inherent complexity and opacity of data-driven models on one hand and the widespread scepticism of the public towards AI [2]. Decision makers and the affected public need trust in forecasts in order to take proposed actions. These challenges provoke questions on how to embed ethical principles in the development of data-driven forecasting, to ensure its acceptance and thus effectiveness:

The quality of applications in weather and climate forecasting is of outstanding importance due to their significant impact on society. Reliability and robustness are critical as these systems must provide accurate and consistent predictions to effectively support various sectors. Transparency is equally crucial since users need to understand how models arrive at their conclusions to trust and act upon their forecasts. From these ethical principles follow measures to ensure adequate risk management, data quality, explainability, and security.

National weather services and intergovernmental organisations like the European Centre for Medium-Range Weather Forecasts (ECMWF) bear elevated accountability towards national governments, which have the vital role in their mandate to issue warnings to the public and plan local protection measures. Long standing efforts in operational meteorology to ensure adherence to standards and foster trust provide suitable ground for implementing the necessary measures to develop and maintain trustworthy AI systems. What distinguishes the ECMWF and national weather forecasting services is the fully transparent and open approach they take in addressing these topics.

This whitepaper investigates the ethical implications of data-driven weather prediction, with a particular focus on the DestinE initiative. It follows the below structure:

- **Application of AI in Weather and Climate Forecasting (Section 3):** This section covers the evolution and current applications of ML in forecasting. It showcases AI-related activities in DestinE and novel models like ECMWF's Artificial Intelligence Forecasting System (AIFS), highlighting the transformative impact on and differences to traditional NWP methods.
- **Ethical considerations (Section 4):** This section discusses ethical considerations that arise from the increased use of ML in meteorology, within weather and climate

modelling itself as well as considerations in relation to systems that are developed downstream from pure weather forecasts. This includes exploring potential risks and impacts associated with downstream applications inspired by use cases implemented for the DestinE digital twins. It illustrates use cases for short to medium-range[2] forecasting, such as disaster warning systems and wind farm management, as well as climate modelling use cases like investment decisions for renewable energy infrastructure and strategies for polar navigation and fishing.
- **Approaches to address ethical concerns (Section 5):** The final section examines approaches to ensure the trustworthiness and adoption of data-driven forecasting systems, addressing both the importance of openness and transparency as well as general frameworks for trustworthy AI and their adoption as practical guidelines for the development of ethical ML systems.

We will conclude with an outlook in Section 6. A more detailed discussion of topics such as robustness and replicability are covered in the whitepaper titled "Robustness and reproducibility for data-driven weather forecasting" [a] and topics such as explainability and resilience towards malicious influence are covered in the whitepaper titles "Explainability and Adversarial Vulnerability in AI-Based Weather Forecasting Systems" [b].

[2] Medium-range is defined as weather forecasting between 3 to 15 days ahead.

3 Increased use of AI in the meteorological community

Like in many other areas, the dramatic advances of AI in recent years had a big impact within the weather and climate community. This section will give a brief overview of recent developments in context of this discussion.

3.1 An overview of developments

Historically, NWP systems like the ECMWF's Integrated Forecasting System (IFS) have been effective in describing atmospheric physics, enabling accurate medium-range forecasts. Over the decades, these traditional models have seen significant improvements in accuracy, with advances in computational power and data assimilation techniques enhancing their predictive capabilities [1], [3].

However, traditional NWP systems are based on solving nonlinear partial differential equations that describe physical processes, such as the Navier-Stokes equations, over fine spatial grids with high time resolution, a computationally intensive approach that relies on high-performance supercomputers. Probabilistic forecasts are created by running ensembles of many such deterministic simulations with slightly perturbed initial conditions and model parameters [4].

In contrast, ML-based models offer a computational advantage at inference time, as they bypass the need to iteratively solve physical equations. Instead, ML approximates physical processes like atmospheric dynamics by learning patterns directly from data [1].

Inspired by the successes of artificial neural networks in diverse domains, the last years have shown rapid progress in the development of data-driven models for weather forecasting, based on deep learning architectures such as convolutional neural networks (CNNs) [5], [6], transformer-based models [7], and graph neural networks (GNNs) [8].

Starting from simple CNNs that predicted only few atmospheric variables, the complexity of models and completeness of variables increased rapidly since 2022, with multiple influential research proposing methods that are approaching or surpassing the performance of the leading traditional NWP systems like the IFS [1], first under specific circumstances and only for certain metrics [8], [9], then for a broad range of variables.

These advancements were enabled principally by the free availability of large-scale, high-quality datasets like the ERA5 reanalysis dataset by the EU's Copernicus Climate Change service [5], [10]. These developments are driven primarily by large corporations: Pangu-Weather [9] by Huawei, GraphCast [11] and GenCast [12] by Google Deepmind, FourCastNet [13] by NVIDIA, Aurora [14] by Microsoft, but ECMWF is also at the forefront, having put the first AI-based forecasts system, the AIFS (Artificial Intelligence Forecasting System), in operation alongside the IFS in February 2025.

Because the resolution on which these models perform well is typically tied closely to the resolution of the reanalysis dataset used for training, their resolution is still much lower than current operational NWP models. Microsoft's Aurora model is the first model that surpasses IFS's skill at a resolution of 11km close to IFS's 9km, combining extensive pre-training on heterogeneous datasets (climate simulations, reanalysis data and operational forecasts) with differing resolutions and fine-tuning on multiple tasks [14].

Taking the use of ML even further, another new approach tries to learn the relation between observations and target atmospheric variables directly without relying on prior data assimilation. These models are trained on historical time series of satellite and conventional observations instead of reanalysis datasets. Two recent examples of this new fully data-driven approach are AI-DOP and GraphDOP, both developed by ECMWF [16], [17]. While showing first promising results this area will need more research conducted under the Horizon Europe Weather Generator project.

Besides the core function of forecasting meteorological variables, ML models have begun to replace or complement traditional systems in key areas like data assimilation [18] or the parameterization of sub-grid scale processes, such as cloud formation, turbulence, and radiation, to avoid the previously required time-consuming and error-prone manual tuning by researchers [5]. Uncertainty quantification of traditional NWP ensembles [19], and downscaling as well as other post-processing steps are another supplementary use of ML [20].

Although not a domain-specific application, the recent advances in natural language processing and large language models will further revolutionise how users interact with forecast data as this allows the easier fusion with other data sources and allow for personalised insights into forecasts. Various chatbots within the community are in development and also DestinE itself is in the process of releasing its own chatbot supporting the use and interaction with DestinE services and data sources. Future developments might include support of mainstream public chatbots, rather than supporting own ones, which will bring its own ethical challenges.

The orders of magnitude higher computational efficiency of ML-based models at forecast time is at the heart of their appeal, as it relaxes the major constraint on traditional NWP systems. It allows to increase forecast resolution while still providing timely forecasts keeping constant or increasing the number of ensemble members. The increased computational efficiency also results in lower energy consumption during operational use.

For climate projections, the ability to rapidly run multiple scenarios in the same amount of time is a key advantage that allows researchers to explore a wider range of potential outcomes under different climate interventions or policy pathways. This capability improves the decision-making process by providing more rounded insights into long-term trends and uncertainties [21], enhancing our capacity to prepare for and respond to weather events and climate change [22]. This ability to run many scenarios can also benefit disaster preparedness [23].

AI models also demonstrate potentially greater accuracy and improved robustness in handling noisy or incomplete data, particularly in short-term forecasting and post-processing tasks. However, in scenarios involving highly irregular datasets with significant uncertainty, traditional NWP ensemble systems still provide more reliable long-range predictions [13].

3.2 Validation of Classical versus AI Models

The validation of forecasting models, whether classical or data-driven, plays a critical role in establishing their reliability and operational viability, directly impacting their trustworthiness. Classical NWP systems, such as IFS, have relied on rigorous methods developed over decades to ensure continuous excellence and confidence building. Observational data from weather stations, satellites, and radar systems provide the baseline for comparing predicted outcomes with real-world measurements. Forecast skill is assessed by various metrics, while iterative bias correction techniques ensure that forecasts improve over time [24]. Models also undergo case-specific evaluations, such as analysing the predicted versus observed tracks of hurricanes or other extreme events [25].

Data-driven models, while leveraging similar evaluation metrics, require adaptations to account for their nature. A challenge lies in the smoothness of ML-based forecasts, as data-driven models trained to maximise accuracy can produce outputs that lack the spatial and temporal coherence typically observed in classical systems. One approach to analyse the physical realism of model forecasts is to use power spectra, where energy levels in the forecast at different scales are compared to a realistic baseline [26].

The ensemble-based approach of classical NWP systems captures a wide range of potential outcomes, making them more robust for extended forecasts [27], [28]. While earlier developments of data-driven models often focussed on singular deterministic forecasts, recent models have also shown the capacity to maintain a well-spread ensemble with competitive performance and decreased smoothness [12], [15].

Reanalysis datasets, such as ERA5 serve as benchmarks for validating AI predictions, offering consistency and comparability with established standards. High-resolution forecast datasets, complement these benchmarks by providing additional validation points for specific scenarios [11]. Advancements in validation models and frameworks, such as cross-disciplinary validation techniques from meteorology and data science, are also emerging to enhance model reliability [29].

As different stakeholders are interested in different facets of weather forecasting there is no single metric that can be used to evaluate the quality of a forecasting model, and different models will lead for specific variables and use cases. Nonetheless, standardised benchmarks like the community-developed WeatherBench [30] and its successors serve as a way to allow an easily reproducible comparison of various scores between data-driven models and NWP baselines taking into account the probabilistic nature of weather forecasting. Another approach through open competitions is trialled with the support of DestinE. The AI Weather Quest[3] started in summer 2025 and concentrates on the sub-seasonal to seasonal (S2S) domain and tries to establish methods for better intercomparisons of global AI-based weather models.

3.3 The DestinE Initiative

Destination Earth (DestinE) is a flagship initiative of the European Commission to develop highly accurate digital models of the Earth, known as a digital twin. The initiative aims to monitor, simulate, and predict the interaction between natural phenomena and human activities, supporting the green transformation and contributing to achieving the objectives of the European Green Deal and Digital Strategy [31].

The DestinE initiative represents an ambitious effort to leverage European assets in high-performance computing (HPC) and AI to develop novel capabilities for better informed decision making. By integrating diverse datasets and enabling real-time analysis, DestinE seeks to improve global environmental understanding, support decision-making in disaster response and climate adaptation, and foster sustainable development. The planned km-scale resolution also captures phenomena on a scale below that of traditional grid cells, allowing for decision support on the local level [32], [33], [34].

For a discussion of AI-related activities in DestinE it is important to understand both the interplay of DestinE components and the dependencies between the DTs and other ECMWF activities (Figure 1).

[3] https://aiweatherquest.ecmwf.int

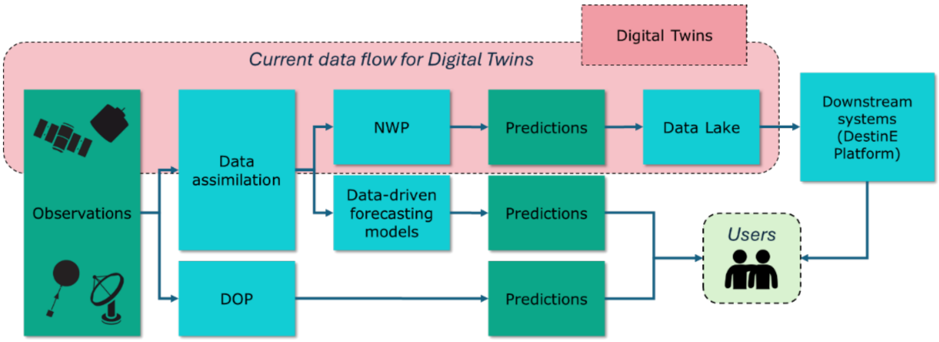

Figure 1: AI-related Activities in and around the DestinE Initiative

Central to DestinE are two high-priority Digital Twins (DTs) – virtual replicas of Earth's systems designed to simulate and predict environmental changes with unprecedented precision. The Digital Twin on Weather-Induced Extremes focuses on detailed simulations of Earth system processes, offering short- to medium-range weather forecasts and insights into extreme events like cyclones, floods, and heatwaves. Meanwhile, the Digital Twin on Climate Change Adaptation enables long-term simulations, helping assess climate risks, test policy interventions, and support sustainable development strategies [34]. At the heart of the two DTs are well established European physics-based NWP models with a limited number of ML components. For the extreme events DT this is ECMWF's Integrated Forecast System (IFS) and for the Climate DT beside IFS variants the German ICON model.

AI serves as a cornerstone of DestinE and its flanking efforts at ECMWF. Currently, the main part of the prediction relies on traditional forecasts based on descriptions of physical dynamics in the form of differential equations, yet minor AI-based components are used or planned to be used in the future, e.g. in uncertainty quantification. Additionally, the recent AIFS model, which is not currently a part of the DTs, is the first fully operational mainly ML-based weather forecasting model by ECMWF. Current work also explores how ML models can be leveraged to approximate ensemble forecasts and to quantify their uncertainty. This approach not only reduces the computational overhead but also as a result allows for faster generation of forecasts at higher resolutions [35].

Both IFS and AIFS build upon the same observation data and data assimilation system. An alternative approach is pioneered with DOP, a set of prototype ML-based systems undertaking the ambitious task to produce weather forecasts directly from meteorological observations without relying on traditional data assimilation.

The output of the DTs is passed on to the DestinE Data Lake (DEDL), together with data from external data spaces. The DestinE Service Platform (DESP) allows downstream users to interact with the data from the DEDL and implement their own specific use cases on top. Multiple demonstrators have already been implemented in a co-design approach to find a balance between applicability for specific local use cases and generalisability to other users [36].

With its emphasis on responsible AI development, including transparency, accountability, and fairness, and the co-design approach at the heart of the development of use cases, the DestinE initiative aligns with ethical standards while meeting the needs of diverse stakeholders [37].

4 Ethical considerations

In light of the significant potential advantages of data-driven models, it is important to also acknowledge the ethical concerns on the large-scale use of AI:

- **Data availability and representation** – Uneven distribution of weather monitoring networks may result in underrepresentation of certain geographical areas or socioeconomic groups, affecting the accuracy and fairness of data-driven forecasts based directly on observations. For example, insurance data used for impact assessments might be missing in less well-off communities, or the lack of observations in areas like Africa. While this has also been also a challenge for physics-based models, they are by design fairer as the laws of physics apply everywhere equally.
- **Accuracy and public trust** – Inaccurate AI-driven predictions can erode public trust, potentially leading to inadequate preparedness for extreme weather events or unnecessary panic and economic disruption. This is a very similar challenge to what physics-based forecast models faced when they were first introduced, and it takes time to build a general confidence in the skill of forecast models. This is made harder by a general perception of mistrust of AI in the wider public.
- **Transparency and accountability** – The complexity of AI algorithms can make it difficult to understand how predictions are made, raising concerns about the transparency of decision-making processes and the ability to hold systems and their operators accountable for errors or failures [38]. While traditional NWP systems are similarly complex, their grounding in physical laws improves their interpretability and increases trust in their predictions.
- **Costs of AI training** – While inference is often cheap to run compared to physical models, the training process is often very expensive, especially on electricity and required technology. This limits the circle of entities with the ability to train these models to a few powerful providers with the necessary capabilities, and challenges aims of society to limited energy consumption.
- **Algorithmic bias and fairness** – AI models trained on historical data may perpetuate existing biases, potentially leading to disproportionate impacts on vulnerable populations and exacerbating social inequalities. While the progress of AI developments has been very rapid there are still doubts how well these models will be in predicting new extreme weather events. Generative AI models excel in predicting future states which they have already seen in their training data, but more research and validation is required to build confidence on predicting previously unseen events.
- **Multitude of forecasts** – With inference costs low and an increasing number of models freely available, it has never been easier to run own weather and climate forecasts. This can lead to an overwhelming offering of possible forecasts blurring the necessary decision making [39]. Models still depend on a good initial data provision for their training and it is likely that models vary strongly in their quality, but users will have the challenge to choose.

Depending on the degree of reliance on ML, the weight of the above considerations varies. For supplementary use of ML in the forecasting pipeline (e.g. in data assimilation or uncertainty quantification) they are minor, whereas they become more significant for data-driven forecasting models. For models that forecast weather directly from observations without classical data assimilation these considerations carry the largest weight.

4.1 Policy landscape – global, European, national

These considerations are also often interrelated. To give an example, the aforementioned concentration of capabilities is with the same entities that also control search engines, end-user operating systems and cloud infrastructure. This dependence on non-European hyperscalers poses a challenge and potential for conflicts of interest. If AI-generated forecasts bypass official channels for warnings and decision making, it could erode trust in established institutions and lead to fragmented dissemination of critical weather information, subverting the national mandate on issuing warnings to the public and planning local measures. Official channels, such as national meteorological services and global initiatives like the World Meteorological Organization (WMO), play a vital role in ensuring standardized, reliable, and equitable access to forecasts. A lack of alignment with these institutions may result in conflicting data, confusing end-users and reducing the overall effectiveness of weather forecasting systems [40]. Under the leadership of EUMETNET's AI program a new working group on AI Ethics was established at the end of 2025 to further align activities in its 33 member organisations across Europe.

On a broader scale, policy organizations such as the European Commission are committed to advancing AI development within Europe. Legislation led by the EU AI Act is designed to address ethical considerations in the use of AI and to ensure a level playing field in the market which brings certainty to investors and developers. The Act introduces a risk-based framework that classifies many public-sector applications – such as social benefits, law enforcement tools, and access to essential public services – as "high-risk." This classification triggers stringent requirements for fundamental rights impact assessments, transparency, human oversight, and the quality of training data, all aimed at preventing discrimination and protecting individual rights.

The EU AI Act is complemented by the EU Charter of Fundamental Rights and the General Data Protection Regulation (GDPR). These regulations establish rules on privacy, data minimization, automated decision-making, and profiling, which are particularly important when governments deploy AI systems to make or support decisions about individuals. These legislative efforts are under constant review and further harmonization is proposed through the 2025 Omnibus legislation[4].

These legislative initiatives are accompanied by substantial investment programmes, such as those led by the European Commission's AI Office, which support the establishment of a code of practice and provide funding for AI development through GenAI4EU [33] and Invest AI [34]. These efforts create opportunities for innovative AI advancements. Additionally, EuroHPC plays an essential role in broadening access to high-performance computing resources by enabling smaller organizations to train large-scale models that would otherwise be cost-prohibitive. This approach underscores a commitment to equitable AI development consistent with European values.

4.2 Layers of trust

Weather and climate forecasting provide societal benefits that are difficult to measure but when quantified as monetary values by all estimates range in the billions of euros. In a simplified value chain economic values are generated by taking observations of the hydro-meteorological state as input, then through modelling producing forecasts for a future state that are communicated and, finally, influence decision making [41], [42], [43]. In a general sense risks are then connected to (opportunity) losses due to wrong decisions being made, either because the predicted events did not materialise, the predictions were misunderstood, or because the actions suggested by those predictions were not taken (e.g. because of a lack of trust). While the latter is a very important aspect to consider, it often receives less attention compared to the accuracy of the predictions themselves. Building and maintaining trust in weather and climate forecasts is crucial, as it directly impacts how individuals, businesses, and governments respond to the information provided. This trust can be influenced by factors such as the transparency of the forecasting process, the consistency of communication, the observed reliability of past forecasts, and the perceived potential for

[4] https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal

manipulation. Addressing these elements can help mitigate the risks associated with disregarded predictions, in turn boosting the societal and economic benefits derived from accurate weather and climate forecasting.

The trust required for the efficacy of weather forecasting, encompasses the trust of the scientific community, of the public and of providers of potential downstream applications. The ITU-T report on trust provisioning [44] distinguishes between three layers of trust: First, there is trust in physical entities, based on physical measurements or tests (in the case of weather forecasting this would amount to being able to trust sensor data used to inform the predictions).

Further, there is the level of cyber trust, encompassing technical infrastructure and processes to maintain system integrity and availability as well as data protection. Both these aspects further motivate requirements regarding accuracy, robustness and cybersecurity as well as for quality, data and risk management.

Finally, there is the third layer of trust, social trust. This is a subjective layer of trust, depending on how the results and processes are communicated and how the methods themselves are explained rather than only their objective security and validation. While trust of course has an internal component as well (trust in the outcomes of the internal AI systems), putting these systems in production and attracting users and downstream providers inherently extends the question of trust to external stakeholders.

Due to the subjective nature of the trustee-trustor-relationship, how to build social trust depends on the stakeholders involved. For example, transparency of methodologies is particularly relevant for publishing scientific results. Publication of data (as in the DestinE Data Lake) and models and keeping scientific progress open greatly helps to increase transparency and thereby not only gain reach but also trust in the scientific community.

A further consideration on trust are actors that actively try to undermine trust. There are various stakeholders with an intent to challenge the credibility of, e.g. research on the effects of climate mitigation policies, due to the economic impact of climate mitigation measures (e.g. carbon pricing). Also, weather and climate forecasts influence market behaviours and might encourage misrepresentation (e.g. gas prices might rise in expectation of cold winters). Resilience against adversarial attacks as specific vulnerability of AI systems [45] is an important part of quality management in this context. However, there is also a communicative dimension, with malign actors potentially looking to undermine communicative efforts.

This whitepaper will focus on the more direct impact of inaccuracies in the predictions themselves. A whitepaper on explainability will investigate the impact of disregarded predictions and lack of trust more closely and elaborate on transparency of the AI systems from a technical standpoint as well as the connection between transparency and trust and how transparency can help to build and foster trust.

4.3 Considerations in relation to downstream systems

Weather forecasting and climate projections are often not used as a stand-alone information source but instead used in combination with other data sources for downstream applications that serve specific use cases in a plethora of impact sectors. This role carries with it additional considerations and might fall under the terms of the EU AI Act, that would not arise for weather or climate prediction per se. This is particularly true for the DestinE outputs and demonstrated by the already implemented example use cases.

It is therefore important to support and enable downstream application providers and users, to highlight potential risks and encourage the uptake of DT outputs. Downstream providers might rely on information and documentation on the model a downstream application is built upon, for their own governance processes, and without these it likely is arduous to accurately assess the risks of the downstream application.

Regardless of whether the model is data-driven, hybrid or purely physics-based, from a risk management perspective, understanding of the models' workings, strengths and weaknesses and potential failure modes via testing and documentation can greatly facilitate risk management processes of downstream systems. This mirrors the motivation for obligations for upstream model providers mentioned in the EU AI Act.

From a different perspective, transparency is crucial for gaining the trust of potential downstream users and thereby increasing adoption. Through this market mechanism it is beneficial for the models to have quality, risk and data management procedures and governance structures with clear accountability mechanisms and transparent documentation in place to enable easy adoption of produced future results by a broad group of downstream users. Examples of relevant information for downstream providers and users include questions of licensing, acceptable use, intended tasks, testing, data formats, interaction with external hardware and software and integration into an AI system.

To support a thriving community, ECMWF provides extensive documentations on, for instance, tests and model properties that can be used to build reliable applications for various scenarios.

To illustrate the potential impact of inaccuracies or model failures this section will showcase several case studies from the DestinE initiative. The overview of case studies in this section will be structured along 1) the two high-priority twins of DestinE, 2) different trustworthiness dimensions or aspects to highlight potential downstream use cases where risks could emerge.

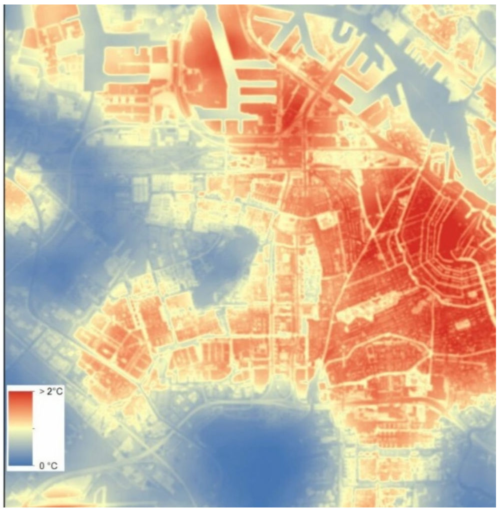

Figure 2: Average daily maximum urban heat island intensity for a section of the city of Amsterdam, highlighting potential street-level impacts of biases on urban planning and climate adaptation. Credit: VITO[5]

4.3.1 Short to Medium-range Forecasting

Inaccurate or unreliable weather forecasts can lead to significant opportunity losses, e.g. farmers might harvest crops at the wrong time, airlines could face unexpected delays, and emergency services might be unprepared for severe weather events. Table 1 illustrates use cases for the Digital Twin on Weather-Induced and Geophysical Extremes both from the public and the commercial domain.

---

[5] https://destine.ecmwf.int/news/destine-for-human-heat-stress-ecmwf-use-case-to-tackle-urban-heat-islands/

Downstream application	Base forecast system	Potential ethical concerns
Flood forecasting	Medium-range	Biased training data can underrepresent rural or marginalized communities, leading to weaker warnings and higher residual risk; unclear model logic makes it hard to assign responsibility when an automated decision leads to missed evacuations; unequal access to advanced AI forecasts may leave low-income regions with poorer protection.
Urban heat stress forecasts	Medium-range	Algorithmic bias can misestimate risks in informal settlements or areas with sparse sensors, worsening health impacts; opaque AI systems reduce public trust and make it difficult to challenge harmful decisions such as closure of cooling facilities; use of fine-grained mobility or location data for forecasting can create privacy risks and enable surveillance.
Support for renewable energy	Medium-range	Grid-optimization models can prioritize reliability for affluent regions while exposing others to more frequent outages; lack of transparency about AI decisions may hide unfair curtailment of certain generators or communities; commercial control over key AI systems and data may create power asymmetries and limit democratic oversight of energy transitions.

Table 1: List of possible downstream applications for extreme weather forecasts and possible ethical concerns.

4.3.2 Climate Modelling

With climate projections, risks and benefits become less direct and harder to quantify.

First, climate change simulations constitute what-if scenarios based on a set of assumptions about, for instance, strategies employed by different governments, population growth, technological advances, and the resulting development of greenhouse gas emissions.

Second, given the inherently unforeseeable nature of the social and economic processes at the basis of these assumptions, the complexity of the Earth's system itself, and the much longer timeframe the uncertainty in those simulations is much larger. This implies that the climate projections do not forecast individual weather events but rather indicate trends of certain climate variables such as mean or extreme temperatures, precipitation, ice cover, or frequencies and average intensities of extreme weather events like storms.

As an added caveat, these what-if projections are nonetheless easily misunderstood as similarly accurate and precise as medium-range weather forecasts, especially when they offer similar resolution as is the case for the DT for Climate Change adaptation. It therefore becomes even more important to communicate unambiguously and emphatically the underlying assumptions, as well as the span of uncertainty. As the global climate exhibits highly non-linear behaviour and certain discrete tipping points [46], transparency about the probability of occurrence of those tail events is also crucial.

Based on data from climate projections, different approaches to climate change adaptation might then be taken, i.e. where and what size of dams to build as protection against floods, how to develop farming to adapt to the changing climate. Often this involves large investment decisions with a time horizon of decades which may be difficult or impossible to reverse later. This highlights the importance of the robustness of modelling predictions and public trust in them over long time horizons.

This section will again highlight use cases, based on the Digital Twin for Climate Change adaptation and their risk profile. Whereas most risk considerations are not peculiar to ML-based forecasts, but equally pertinent to traditional NWP simulations, whether data-driven models, having seen only the climate of the past, perform well under a changing climate is an area of active research. Also, for the use cases in this section, the frequency and intensity of extreme weather events is of relevance. As discussed in Section 3, the smoothing of forecasts by data-driven methods could potentially lead to fewer simulated extreme events, underestimating their occurrence. If these biases and uncertainties remained unconsidered, this could lead to a distortion of decisions.

While the quality of the forecasts is, of course, important it is essential to recognize that even if the assumptions underlying them turn out to be invalid, decisions based on climate projections are nearly always better than without them. The risks associated with the AI systems that are mentioned here are then not relative to a case where no system is in place but rather compared to the fiction of a system with perfect foresight.

Downstream application	**Base forecast system**	**Potential ethical concerns**
Renewable energy investments – planning of wind farms or solar parks	Climate scenarios	Less data from remote or economically disadvantaged regions can lead to unequal investment opportunities; models may not adequately consider the social and environmental impacts on local communities; lack of transparency can cause mistrust of stakeholders; high training costs can potentially lead to monopolies in wind farm investments.
Urban planning	Climate scenarios	Algorithmic bias can misestimate risks of climate change in informal settlements or with sparsely monitored regions, leading to unfair allocation of climate adaptation measures; black-box nature of AI systems makes it difficult to challenge these decisions.
Flood risk estimation	Climate scenarios	Sparse data on precipitation in low-income regions may lead to underestimation of flood risk and underinvestment in protection measures, alternatively overestimation may lead to overpricing of insurance or insufficient coverage.

Table 2: List of possible downstream applications of climate change scenarios and possible ethical concerns.

5 Approaches to address ethical concerns

Following the discussion of ethical considerations due to the increased use of AI in meteorological applications in Section 4, this section will present approaches to address these ethical concerns, and motivate benefits of requirements such as quality and risk management. The DestinE initiative and development of data-driven forecasting models at ECMWF serve as examples for this examination.

As illustrated above, there are many (potential) applications with varying risks that can build on top of the DT or data-driven forecasting systems (e.g. AIFS), giving considerable importance to the ethical concerns raised in Section 4. While many approaches exist to address them from a technical perspective, for example hybrid systems of AI and classical forecasting which try to compensate the weaknesses of one another [47], [48] as well as methods for interpretable ML [49], [50], a detailed discussion of these aspects is delegated to other parts of this series of whitepapers. Instead, communication and collaboration practices in the meteorology community are first examined to find the aforementioned aspects of documentation, transparency and collaboration along the supply chain often fulfilled by open and transparent design.

5.1 Openness in Approaches and Work

Transparency has long been a cornerstone of scientific progress in meteorology. Classical NWP models published their methodologies and results extensively in scientific literature, enabling scrutiny, reproducibility, and collaboration across the research community. Past observations and forecasts data sets are openly available to allow scrutiny and research in new methodologies. This openness established trust in their forecasts and accelerated advancements in the field. As ML-based approaches become prevalent, maintaining a comparable level of transparency and literacy with users is critical. Given the complexity and opacity of many AI models, openly sharing their methodologies, datasets, and performance metrics ensures that advancements are rigorous, reproducible, and widely accessible [51], [52]. This mirrors requirements set forth by the EU AI Act regarding AI literacy and allows users to properly interpret the systems' output and understand the rationale behind their predictions, which is a prerequisite for effective human oversight.

ECMWF exemplifies this openness in its approach to research and development and takes the obligation to literacy of its users very serious. Through platforms such as the AIFS blog, ECMWF provides regular updates on its progress in AI-driven weather forecasting [54]. This transparency fosters collaboration across institutions and facilitates critical evaluation, ensuring that shared methodologies, data, and findings undergo rigorous assessment and refinement, driving collective advancements in meteorological science. By openly communicating their challenges and breakthroughs, ECMWF not only builds trust within the scientific community but also accelerates the development of more accurate and reliable forecasting models. Moreover, outreach activities to stakeholders und policy makers keeps these key public figures appraised of the importance of ethical AI development and increases AI literacy among them.

In addition, education activities provided as part of the DestinE initiative and by ECMWF as an entrusted entity are essential for contributing to AI literacy in the meteorological community. Practical guidelines for developers and practitioners form a tool to verify that all necessary ethical considerations in the development of AI systems have been addressed. In combination with open and free training on AI (and its responsible use) in meteorology these contribute to AI literacy in the field.

Open collaboration and community are the foundation for innovation in weather forecasting, supported by open-source platforms, open frameworks, and competitive challenges. Open-source repositories on platforms like GitHub and Hugging Face provide accessible tools and frameworks for AI development, empowering researchers, institutions, and businesses to rapidly prototype and test new models. These platforms enable smaller organizations to contribute without requiring extensive infrastructure, fostering shared learning and efficiency across the field [38].

Open frameworks further enhance collaboration by ensuring standardization and reproducibility in model development. They streamline deployment and enable consistent benchmarking of results. Developments like those by ECMWF, highlighted through the AIFS blog and a series of webinars, demonstrate how standardized approaches reduce duplication of work and improve the efficiency of model refinement. Reproducibility ensures trust in results and facilitates broader adoption of AI tools in weather prediction, creating a solid basis for future progress [52]. As one example, the Anemoi framework developed by ECMWF in cooperation with multiple national weather services embodies this collaborative ethos, offering a structured environment for creating ML weather forecasting systems. By providing key building blocks and traceability of data and processing choices across the full AI workflow - from data preparation over training to deployment - it offers an open framework reducing not only the complexity in adopting current data-driven models for meteorological organisations but allows good practices to be shared at a large scale.

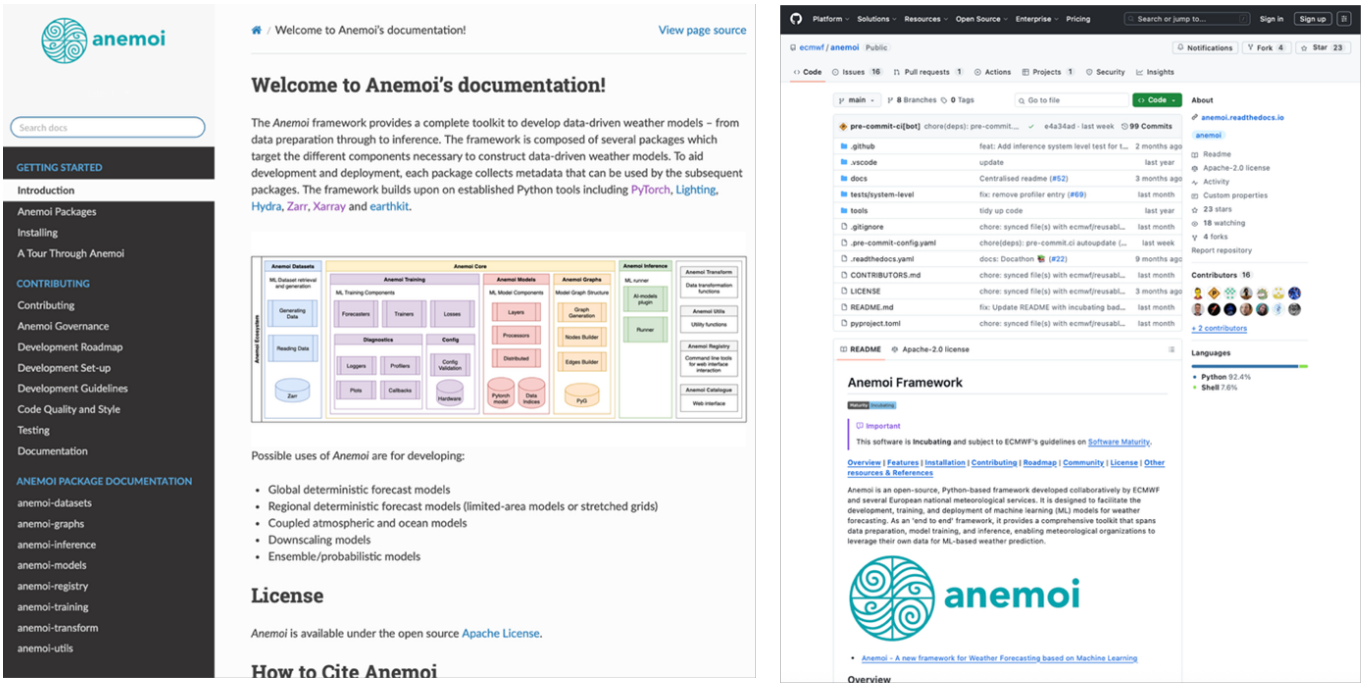

Figure 3: The open approach by the Anemoi framework on public platforms enables the whole community to engage, contribute and build trust in the tools.

Open competitions complement these efforts by driving innovation and engaging diverse contributors from academia, industry, and beyond. By presenting challenges to solve real-world problems, competitions act as a testing ground for novel techniques and provide a platform for validating models under standardized conditions. Incentives such as funding and recognition motivate participation, bridging the gap between theoretical advancements and practical implementation. These events foster collaboration, inspire new applications of AI in weather forecasting, and establish an ecosystem of acceptance and engagement within the field [52].

Together, these open collaboration initiatives enable institutions, businesses, and society to harness AI in meteorology, advancing the science while ensuring its accessibility and reliability.

5.2 General Frameworks for Trustworthy AI

There are many reasons to use a structured approach to include and operationalize ethical concerns in AI development – maintaining and assessing the quality and trustworthiness of the system signalling trustworthiness by including transparent documentation of assessments for downstream providers and making sure that the consideration of ethical aspects doesn't depend on single developers but is contained in underlying processes. For example, examination, test and validation procedures are common practice. In the domain of weather forecasting, models are typically benchmarked on previous weather data as well as challenging data, while undergoing more extensive robustness testing prior to deployment. After deployment, models in use are continuously monitored (based on past predictions) to detect potential errors as well as risks of data or model drift. Such concrete technical aspects will be discussed in more detail in later parts of this whitepaper series, for the remainder of this section the focus will be on the operationalization and inclusion of ethics and trustworthiness in a structured development process.

A common approach to structured incorporation of ethical concerns into the development process of AI systems is to look at frameworks for trustworthy and ethical AI. These are most often designed to be generally applicable, regardless of model architectures and application areas.

Frameworks typically classify risks of AI systems in multiple dimensions such as reliability or fairness. Which risks are particularly relevant for a given AI application then depends on the application context and intended use of the AI system. For an AI system classifying applicants by their suitability for a given job offering it might for example be expected to put a bigger focus on questions of fairness compared to an AI system classifying products in a warehouse. Following the approach of the Fraunhofer AI Assessment Catalog, this step of identifying risks given the context of application of an AI system falls under the term "protection needs analysis".

The Fraunhofer AI Assessment Catalog lists risks and measures along the lifecycle stages of data, development and operation and distinguishes six trustworthiness dimensions: reliability, safety & security, fairness, data protection, transparency and human autonomy & control. Following a protection needs analysis in this dimension intended to capture where measures need to be taken, metrics and goals are defined to operationalize aspects of the trustworthiness dimension.

Protection needs analysis and definition of metrics are naturally highly dependent on the application context and intended use of an AI system as AI systems have vastly different risks and suitable metrics vary between modalities and application domains.

Following this definition of metrics and thresholds, appropriate mitigation measures can be taken to achieve the targets. In the case of fairness for example, upon finding that a model does not perform in a given fairness metric, checking for bias in training data and balancing the training dataset, if necessary by including synthetic data, could lead to an improvement in these metrics.

After mitigating, one can then compare the targets set earlier to the achieved results and aggregate across the trustworthiness dimensions to arrive at a measure for trustworthiness of the AI system as a whole and can therefore also effectively argue the trustworthiness of the AI system.

The structure of such frameworks can vary however, and it is valuable to take different frameworks into account. Other possible dimensions contained in such frameworks for ethical AI are for example sustainability or accountability [58]. While these topics are certainly interesting to examine for the use case at hand, they are often less related to the everyday processes of developers and have therefore been omitted in the creation of the guidelines.

As a part of this project, a protection needs analysis was performed for the (broad) use case of weather and climate forecasting. The outcome of this analysis is visualized in Figure 4**Error! Reference source not found.** below. Out of the six dimensions, four were deemed relevant. Note that this does not necessarily apply for downstream use cases. In the case of catastrophe prevention for example, keeping an element of human supervision would be crucial. Similarly, data protection could play a role when considering personalized weather apps using location data. On the other hand, weather data itself is often publicly available and therefore not in need of specific measures to prevent access to it. Measures to prevent manipulations of training or inference data are relevant but fall under safety & security.

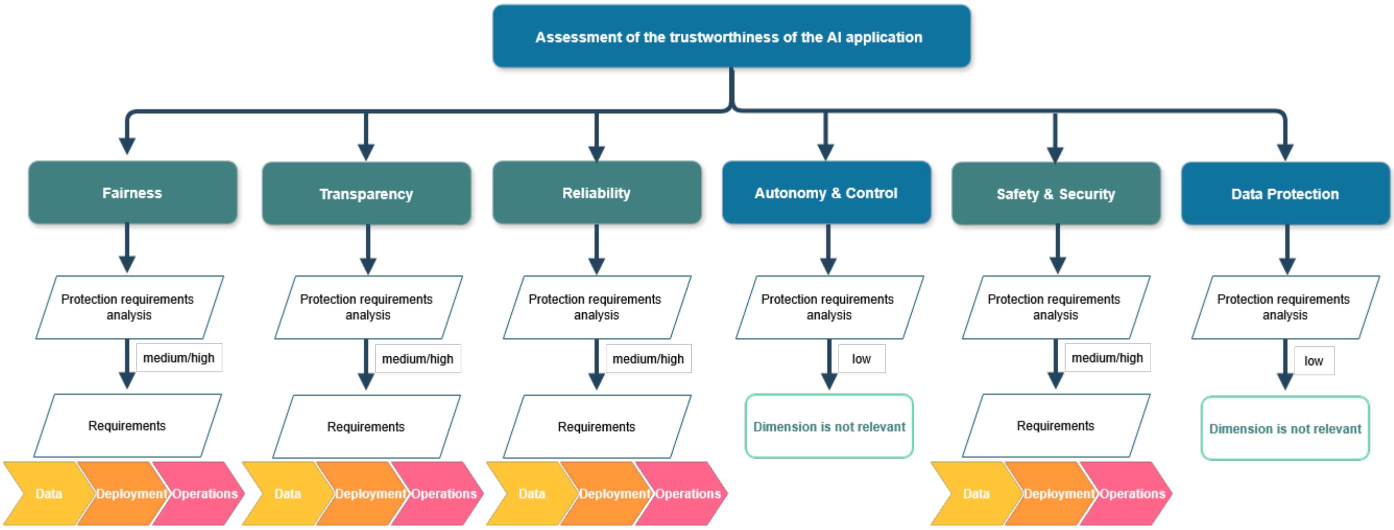

Figure 4: Relevant dimensions for the trustworthiness assessment

As a result, a set of guidelines [c] were developed, suitable for developers searching for guidance on how to approach ethical concerns during the development process of an AI system. The guidelines offer potential measures along a simplified AI lifecycle – from data over the model to deployment and operations.

5.3 Practical Guidelines

The guidelines take into account the four identified trustworthiness dimensions of relevance and are structured to give potential metrics and mitigation measures along a simplified AI lifecycle containing the lifecycle stages "data", "deployment" and "operation".

In a first step, metrics and measures from the Fraunhofer AI Assessment Catalog [56] were reviewed and formed the basis of the guidelines. The structure was simplified by integrating metrics and measures and introducing a new sub-structure below trustworthiness dimensions to replace the risk areas of the AI Assessment Catalogue. This allows to focus on relevant aspects of the given use case and adapt the structure accordingly.

In collaboration with developers from ECMWF, trustworthiness aspects occurring in weather and climate forecasting systems were then identified and integrated into the guidelines.

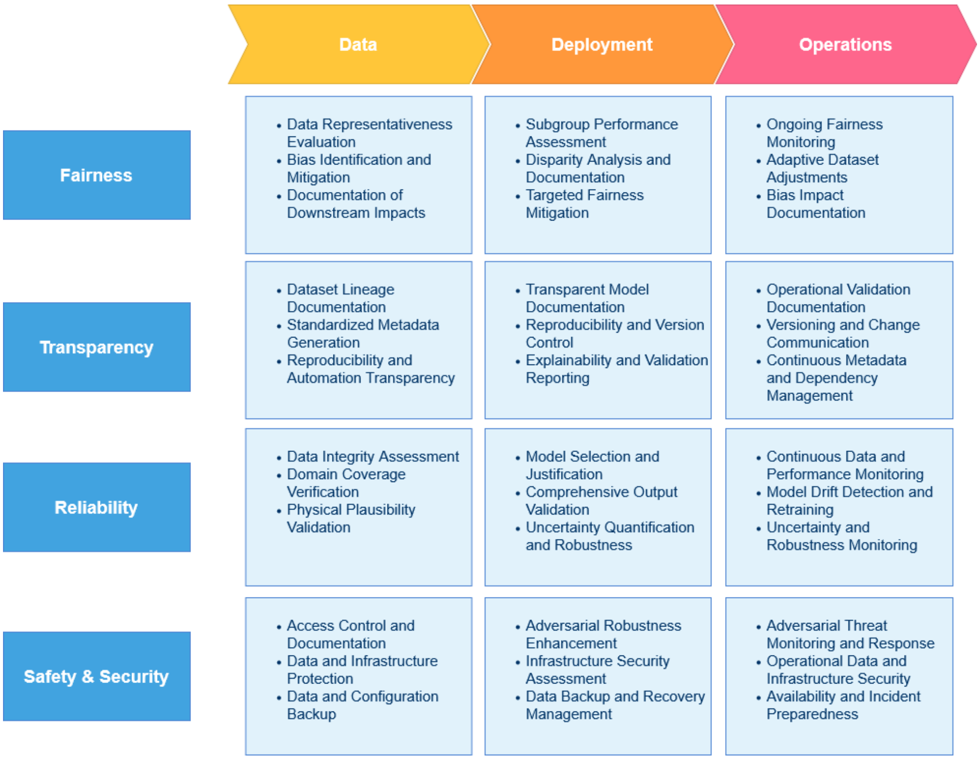

Figure 5: Structure of the practical guidelines forming a matrix of trustworthiness dimensions and lifecycle stages

To enhance clarity and usability, the guidelines were structured by trustworthiness dimensions and lifecycle stages and were further prioritized to allow developers to understand immediately, which requirements are established practice, and which ones might involve questions still being researched.

5.4 Other identified topics

There are topics that were highlighted in Section 4 but could not be investigated in more detail within the scope of our current investigation.

Among other benefits of data-driven forecasts, computational efficiency of data-driven forecasting and uncertainty quantification at time of inference are mentioned. Inversely, their training consumes considerable energy. As a consequence, the total effect of the integration of ML-based methods over the entire lifecycle of the system is less clear. This could serve as a potential starting point for an investigation comparing lifecycle carbon costs of physical versus data-driven systems and canvassing sustainability measures with respect to high-performance compute clusters.

One of the key traits of the current wave of AI is the continuing rapid development of the technologies. Recently, there has been an increasing trend towards agentic AI, i.e. systems that plan, act, and adapt with limited human direction. While meteorological services have specific, high accountability requirements and place great value on human expertise, the field is moving fast and there could be potential applications for autonomous agents acting based on weather forecasts: e.g. automatically orienting wind turbines, releasing weather balloons to fill gaps in observations, disseminating warnings. With this degree of machine autonomy, the ethical considerations from Section 4 particularly regarding human autonomy become more pressing and new mechanisms for effective human oversight would need to be established.

6 Conclusion and Outlook

This whitepaper has explored the ethical implications of the adoption of ML-based systems in weather and climate forecasting. From our discussion a recurring topic has emerged: Weather and climate forecasting serves to enable informed decision-making. For the forecasts to become effective, they must be trusted by decision makers (whether from the administration, commercial entities or the wider public). This whitepaper has touched on several aspects that are necessary to establish trust in data-driven forecasts:

1. The forecasting systems need to be robust, meaning they need to operate reliably under previously unseen input data. Since AI systems learn from past observed data, making them robust against shifts (e.g. changing climate) and imbalances (e.g. sparsely covered regions) in the data remains a challenge. Intricately linked to this robustness is the quantification of how uncertain the predictions are on which decision makers base their decisions. Another prerequisite for robustness is the open and free availability of forecast and training data, allowing the forecasts to be scrutinised. The meteorological community can look back on decades of established validation processes but needs to adapt these to the new paradigm.
2. Additionally, to increase confidence in the predictions, decision makers should be able to understand how they were derived and know about possible biases. The complexity of the ML models makes it impossible to fully interpret the basis for their predictions. However, research into their explainability has made large progress. In line with this, it is vital to support increased AI literacy with developers and users alike, which needs to be reflected in a comprehensive training program.
3. Finally, decision makers need to trust that the predictions were not manipulated. Research has shown that ML models are, in theory, vulnerable to adversarial attacks – techniques where training data or inputs are deliberately manipulated to alter ML models and their predictions. Despite the absence of these attacks in operational environments and the high level of expertise and sustained effort required by potential attackers, the mere potential for such threats can erode stakeholder trust, stressing the necessity for robust security measures.

This series of whitepapers includes two accompanying whitepapers on robustness [a] and explainability [b]. These will highlight differences between physical and ML-based models, the ethical implications and measures to address these aspects in the context of the DestinE initiative.

Looking ahead, the role of AI in weather and climate forecasting is likely to expand, offering new opportunities for innovation, collaboration, and societal benefit. Initiatives like DestinE could significantly impact the advancement of AI in the eyes of the public and regulatory bodies. A large-scale ethically governed AI project can help build trust and set a precedent for future initiatives.

Following best practices of ML development is important from an ethical standpoint resting on European values as well as a pragmatic perspective considering that trust in forecasts is a crucial factor in their effectiveness. ECMWF is therefore taking these prerequisites into account in the development of the DTs and ML-based forecasting systems like AIFS and DOP. To operationalise these in an easily digestible form, ECMWF is working on practical guidelines for involved stakeholders like developers, domain experts, and project managers.

Moreover, it remains important to monitor the evolving legal landscape to promptly adapt to changing requirements. Long standing efforts in operational meteorology to ensure adherence to standards and foster trust in combination with established general frameworks for trustworthy AI already provide a suitable basis for implementing the necessary measures to ensure trustworthiness of the systems.

Looking ahead, the role of AI in weather and climate forecasting is likely to expand, offering new opportunities for innovation, collaboration, and societal benefit. Initiatives like DestinE could significantly impact the advancement of AI in the eyes of the public and regulatory bodies. The example of ML-based weather and climate forecast systems is a powerful use case to challenge and improve the developing guidance on ethical AI use in Europe. A large-scale ethically governed AI project can help build trust and set a precedent for future initiatives.

In conclusion, the integration of AI into weather and climate forecasting offers substantial benefits but also introduces potential risks, particularly when these models are used in downstream applications. By adhering to high standards of robustness, and transparency, and by fostering open communication and collaboration, the field can maximize the societal benefits of these technologies while mitigating potential risks. The future of weather and climate forecasting lies in the responsible and ethical development and integration of AI systems, ensuring that they serve the best interests of society and contribute to a safer, more informed world.

7 References

[a] DE398 Whitepaper – Robustness and reproducibility in data-driven weather forecasting, 02.02.2026

[b] DE398 Whitepaper – Explainability and Adversarial Vulnerability in AI-Based Weather Forecasting Systems, 02.02.2026

[c] DE398 Guidelines – Practical Guidelines on Ethical Machine Learning, 29.01.2026

[1] Z. Ben Bouallègue *et al.*, "The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning–Based Weather Forecasts in an Operational-Like Context," *Bulletin of the American Meteorological Society*, vol. 105, no. 6, pp. E864–E883, Jun. 2024, doi: 10.1175/BAMS-D-23-0162.1.

[2] M. Gerlich, "Perceptions and Acceptance of Artificial Intelligence: A Multi-Dimensional Study," Social Sciences, vol. 12, no. 9, p. 502, Sep. 2023, doi: 10.3390/socsci12090502.

[3] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," Nature, vol. 525, no. 7567, pp. 47–55, Sep. 2015, doi: 10.1038/nature14956.

[4] M. J. Rodwell and T. N. Palmer, "Using numerical weather prediction to assess climate models," Quart J Royal Meteoro Soc, vol. 133, no. 622, pp. 129–146, Jan. 2007, doi: 10.1002/qj.23.

[5] M. G. Schultz et al., "Can deep learning beat numerical weather prediction?," Phil. Trans. R. Soc. A., vol. 379, no. 2194, p. 20200097, Apr. 2021, doi: 10.1098/rsta.2020.0097.

[6] E. Walsh, G. Bessardon, E. Gleeson, and P. Ulmas, "Using machine learning to produce a very high resolution land-cover map for Ireland," Adv. Sci. Res., vol. 18, pp. 65–87, May 2021, doi: 10.5194/asr-18-65-2021.

[7] Y. Wu and W. Xue, "Data-Driven Weather Forecasting and Climate Modeling from the Perspective of Development," Atmosphere, vol. 15, no. 6, p. 689, Jun. 2024, doi: 10.3390/atmos15060689.

[8] R. Keisler, "Forecasting Global Weather with Graph Neural Networks," Feb. 15, 2022, arXiv: arXiv:2202.07575. doi: 10.48550/arXiv.2202.07575.

[9] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, "Accurate medium-range global weather forecasting with 3D neural networks," Nature, vol. 619, no. 7970, pp. 533–538, Jul. 2023, doi: 10.1038/s41586-023-06185-3.

[10] B. Bochenek and Z. Ustrnul, "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives," Atmosphere, vol. 13, no. 2, p. 180, Jan. 2022, doi: 10.3390/atmos13020180.

[11] R. Lam et al., "Learning skillful medium-range global weather forecasting," Science, vol. 382, no. 6677, pp. 1416–1421, Dec. 2023, doi: 10.1126/science.adi2336.

[12] I. Price et al., "GenCast: Diffusion-based ensemble forecasting for medium-range weather," May 01, 2024, arXiv: arXiv:2312.15796. doi: 10.48550/arXiv.2312.15796.

[13] M. Adrian, D. Sanz-Alonso, and R. Willett, "Data Assimilation with Machine Learning Surrogate Models: A Case Study with FourCastNet," 2024, arXiv. doi: 10.48550/ARXIV.2405.13180.

[14] C. Bodnar et al., "A Foundation Model for the Earth System," Nov. 21, 2024, arXiv: arXiv:2405.13063. doi: 10.48550/arXiv.2405.13063.

[15] S. Lang et al., "AIFS-CRPS: Ensemble forecasting using a model trained with a loss function based on the Continuous Ranked Probability Score," Dec. 20, 2024, arXiv: arXiv:2412.15832. doi: 10.48550/arXiv.2412.15832.

[16] M. Alexe et al., "GraphDOP: Towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations," Dec. 20, 2024, arXiv: arXiv:2412.15687. doi: 10.48550/arXiv.2412.15687.

[17] A. McNally et al., "Data driven weather forecasts trained and initialised directly from observations," Jul. 22, 2024, arXiv: arXiv:2407.15586. doi: 10.48550/arXiv.2407.15586.

[18] J. D. Keller and R. Potthast, "AI-based data assimilation: Learning the functional of analysis estimation," 2024, arXiv. doi: 10.48550/ARXIV.2406.00390.

[19] P. Grönquist et al., "Deep learning for post-processing ensemble weather forecasts," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 379, no. 2194, p. 20200092, Feb. 2021, doi: 10.1098/rsta.2020.0092.
[20] N. Rampal, P. B. Gibson, S. Sherwood, G. Abramowitz, and S. Hobeichi, "A Reliable Generative Adversarial Network Approach for Climate Downscaling and Weather Generation," J Adv Model Earth Syst, vol. 17, no. 1, p. e2024MS004668, Jan. 2025, doi: 10.1029/2024MS004668.
[21] S. Materia et al., "Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives," WIREs Climate Change, vol. 15, no. 6, p. e914, Nov. 2024, doi: 10.1002/wcc.914.
[22] X. X. Zhu et al., "On the Foundations of Earth and Climate Foundation Models," May 07, 2024, arXiv: arXiv:2405.04285. doi: 10.48550/arXiv.2405.04285.
[23] K. Feng, D. Xi, W. Ma, C. Wang, Y. Li, and X. Chen, "Potential Paradigm Shift in Hazard Risk Management: AI-Based Weather Forecast for Tropical Cyclone Hazards," Apr. 29, 2024, arXiv: arXiv:2404.18440. doi: 10.48550/arXiv.2404.18440.
[24] R. Buizza et al., "The development and evaluation process followed at ECMWF to upgrade the Integrated Forecasting System (IFS)," ECMWF Technical Memoranda, no. 829. ECMWF, Oct. 2018. doi: 10.21957/xzopnhty9.
[25] E. Zsótér, "Recent developments in extreme weather forecasting," ECMWF Newsletter, 2006, doi: 10.21957/KL9821HNC7.
[26] Zied Ben Bouallègue and ECMWF AIFS Team, "Accuracy versus activity," ECMWF Newsletter, 2024, doi: 10.21957/8B50609A0F.
[27] ECMWF, "Quantifying forecast uncertainty," Quantifying forecast uncertainty. [Online]. Available: https://www.ecmwf.int/en/research/modelling-and-prediction/quantifying-forecast-uncertainty
[28] ECMWF, "Ensemble weather forecasting." Accessed: Dec. 06, 2024. [Online]. Available: https://www.ecmwf.int/sites/default/files/medialibrary/2017-03/ecmwf-fact-sheet-ensemble-forecasting.pdf
[29] T. C. Pagano et al., "Challenges of Operational Weather Forecast Verification and Evaluation," Bulletin of the American Meteorological Society, vol. 105, no. 4, pp. E789–E802, Apr. 2024, doi: 10.1175/BAMS-D-22-0257.1.
[30] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, "WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting," Journal of Advances in Modeling Earth Systems, vol. 12, no. 11, p. e2020MS002203, 2020, doi: 10.1029/2020MS002203.
[31] "Destination Earth | Shaping Europe's digital future." Accessed: Feb. 18, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/destination-earth
[32] ECMWF, "Destination Earth: initiative explanation." Accessed: Jun. 12, 2024. [Online]. Available: https://www.ecmwf.int/en/about/what-we-do/environmental-services-and-future-vision/destination-earth
[33] ECMWF, "Disaster Risk Mitigation & Climate Adaptation," Disaster Risk Mitigation & Climate Adaptation. Accessed: Jun. 12, 2024. [Online]. Available: https://destine.ecmwf.int/use-case/destine-use-case-disaster-risk-mitigation-climate-adaptation/
[34] ECMWF, "Destination Earth's digital twins and Digital Twin Engine – state of play." Accessed: Jun. 12, 2024. [Online]. Available: https://www.ecmwf.int/en/newsletter/180/earth-system-science/destination-earths-digital-twins-and-digital-twin-engine
[35] M. Chantry, H. Christensen, P. Dueben, and T. Palmer, "Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI," Phil. Trans. R. Soc. A., vol. 379, no. 2194, p. 20200083, Apr. 2021, doi: 10.1098/rsta.2020.0083.

[36] J. Hoffmann, P. Bauer, I. Sandu, N. Wedi, T. Geenen, and D. Thiemert, "Destination Earth – A digital twin in support of climate services," Climate Services, vol. 30, p. 100394, Apr. 2023, doi: 10.1016/j.cliser.2023.100394.
[37] ECMWF, "Phase Two of Destination Earth Confirmed." Accessed: Jun. 12, 2024. [Online]. Available: https://stories.ecmwf.int/phase-2/index.html
[38] P. M. Lukacz, "Developing AI for Weather Prediction: : Ethics of Design and Anxieties about Automation at the US Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography," S&TS, Jan. 2024, doi: 10.23987/sts.125741.
[39] G. Phillips-Wren and M. Adya, "Decision making under stress: the role of information overload, time pressure, complexity, and uncertainty," Journal of Decision Systems, vol. 29, no. sup1, pp. 213–225, Aug. 2020, doi: 10.1080/12460125.2020.1768680.
[40] World Meteorological Organization (WMO), "Future of weather and climate forecasting." WMO. Accessed: Dec. 18, 2024. [Online]. Available: https://library.wmo.int/viewer/57328/download?file=1263_WMO_Open_Consultative_Platform_White_Paper_en.pdf&type=pdf&navigator=1
[41] "Weather-Water-Climate Value Chain(s): Giving VOICE to the Characterization of the Economic Benefits of Hydro-Met Services and Products," American Meteorological Society. Accessed: Jan. 17, 2025. [Online]. Available: https://www.ametsoc.org/index.cfm/ams/policy/studies-analysis/weather-water-climate-value-chain-s-giving-voice-to-the-characterization-of-the-economic-benefits-of-hydro-met-services-and-products/
[42] F. Pappenberger, H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson, and J. Thielen, "The monetary benefit of early flood warnings in Europe," Environmental Science & Policy, vol. 51, pp. 278–291, Aug. 2015, doi: 10.1016/j.envsci.2015.04.016.
[43] "Charting a Course for Sustainable Hydrological and Meteorological Observation Networks in Developing Countries," 2022, Accessed: Jan. 17, 2025. [Online]. Available: https://hdl.handle.net/10986/38071
[44] M. Roshanaei and Q. Duan, "International Telecommunication Union Standardization for Trust Provisioning in Information, Communication and Technology Infrastructure toward Achieving United Nation's Sustainable Development Goals," JCC, vol. 09, no. 10, pp. 44–59, 2021, doi: 10.4236/jcc.2021.910004.
[45] C. Szegedy et al., "Intriguing properties of neural networks," in International Conference on Learning Representations, 2014. [Online]. Available: http://arxiv.org/abs/1312.6199
[46] T. M. Lenton et al., "Climate tipping points — too risky to bet against," Nature, vol. 575, no. 7784, pp. 592–595, Nov. 2019, doi: 10.1038/d41586-019-03595-0.
[47] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next," J Sci Comput, vol. 92, no. 3, p. 88, Sep. 2022, doi: 10.1007/s10915-022-01939-z.
[48] D. Kochkov et al., "Neural General Circulation Models for Weather and Climate," Nature, vol. 632, no. 8027, pp. 1060–1066, Aug. 2024, doi: 10.1038/s41586-024-07744-y.
[49] S. Kim et al., "Explainable AI-Based Interface System for Weather Forecasting Model," in HCI International 2023 – Late Breaking Papers, vol. 14059, H. Degen, S. Ntoa, and A. Moallem, Eds., in Lecture Notes in Computer Science, vol. 14059. , Cham: Springer Nature Switzerland, 2023, pp. 101–119. doi: 10.1007/978-3-031-48057-7_7.
[50] R. Yang et al., "Interpretable machine learning for weather and climate prediction: A review," Atmospheric Environment, vol. 338, p. 120797, Dec. 2024, doi: 10.1016/j.atmosenv.2024.120797.
[51] D. Schuster and M. Friedman, "AMS Publications Support for Open, Transparent, and Equitable Research," Bulletin of the American Meteorological Society, vol. 104, no. 11, pp. 899–900, Nov. 2023, doi: 10.1175/BAMS-D-23-0243.1.
[52] G. Brunet et al., "Advancing Weather and Climate Forecasting for Our Changing World," Bulletin of the American Meteorological Society, vol. 104, no. 4, pp. E909–E927, Apr. 2023, doi: 10.1175/BAMS-D-21-0262.1.

[53] ECMWF, “AIFS Blog.” Accessed: Dec. 18, 2024. [Online]. Available: https://www.ecmwf.int/en/about/media-centre/aifs-blog
[54] L. Ji, Z. Wang, M. Chen, S. Fan, Y. Wang, and Z. Shen, “How Much Can AI Techniques Improve Surface Air Temperature Forecast? —A Report from AI Challenger 2018 Global Weather Forecast Contest,” J Meteorol Res, vol. 33, no. 5, pp. 989–992, Oct. 2019, doi: 10.1007/s13351-019-9601-0.
[55] P. Herruzo et al., “High-resolution multi-channel weather forecasting – First insights on transfer learning from the Weather4cast Competitions 2021,” in 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA: IEEE, Dec. 2021, pp. 5750–5757. doi: 10.1109/BigData52589.2021.9672063.
[56] M. Poretschkin et al., “Guideline for Trustworthy Artificial Intelligence -- AI Assessment Catalog,” 2023, arXiv. doi: 10.48550/ARXIV.2307.03681.

This document has been produced in the context of the Destination Earth Initiative and relates to tasks entrusted by the European Union to the European Centre for Medium-Range Weather Forecasts implementing part of this Initiative.

This document is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. The European Centre for Medium-Range Weather Forecasts is not liable in respect of this document and gives no warranty for the information provided.

UK: (Headquarters) ECMWF, Shinfield Park, Shinfield Road, Reading, RG2 9AX, UK

Italy: ECMWF, Tecnopolo di Bologna, Via Stalingrado 84/3, 40128 Bologna, Italia

Germany: ECMWF, Robert-Schuman-Platz 3, 53175 Bonn, Deutschland

ecmwf.int/destinE   ecmwf.int