Responsible Artificial Intelligence for Earth Observation

©SHUTTERSTOCK.COM/P.SIRIPAK

Achievable and realistic paths to serve the collective good

PEDRAM GHAMISI, WEIKANG YU, ANDREA MARINONI, CAROLINE M. GEVAERT, CLAUDIO PERSELLO, SIVASAKTHY SELVAKUMARAN, MANUELA GIROTTO, BENJAMIN P. HORTON, PHILIPPE RUFIN, PATRICK HOSTERT, FABIO PACIFICI, AND PETER M. ATKINSON

In the evolving landscape of Earth observation (EO), the fusion of artificial intelligence (AI) with EO holds tremendous promise for unlocking insights into our planet's dynamics [1]. Specifically, the integration of AI techniques for EO (AI4EO) significantly enhances our capacity to distill meaningful information from remotely sensed data [2], which has undergone a substantial surge since 2020, as shown in Figure 1. Nonetheless, the complex nature of Earth systems requires a thorough understanding of AI's inherent limitations and potential consequences, which necessitates the implementation of responsible AI practices in EO. This shapes the particular focus of this article.

Figure 2 shows an overview of the development of AI algorithms over the last decades. The birth of AI as a distinct field is often attributed to a workshop held at Dartmouth College in the summer of 1956. The concept of responsible AI gained significant attention in the late 2010s and early 2020s, with generative adversarial approaches as one of the catalysts. During this period, concerns about AI biases, fairness, transparency, and accountability began to gain prominence, with early influential works, such as "Gender Shades" (2018) [16], highlighting biases in facial recognition technologies. This era also saw the release of ethical guidelines, like IEEE's "Ethically Aligned Design" (2016) [17] and the Association for Computing Machinery's updated Code of Ethics (2018) [18]. In the early 2020s, regulatory efforts intensified with the European Commission's" Ethics Guidelines for Trustworthy AI" (2019) [19] and the OECD's AI Principles (2019) [20]. Major tech companies like Google and Microsoft established dedicated responsible AI teams and initiatives. The concept of responsible AI continued to evolve, integrating into AI development pipelines and emphasizing both data quality and model innovation. Standardization efforts by organizations like the International Organization for Standardization and educational initiatives further promoted responsible AI practices, addressing ethical considerations, such as fairness, transparency, and bias mitigation. In our opinion, and as indicated by the light orange color in Figure 2, we have transitioned from an era of singular focus on either model-centric or data-centric approaches to a new era characterized by balanced

Digital Object Identifier 10.1109/MGRS.2025.3529726
Date of publication 21 February 2025; date of current version 19 September 2025.

data–model approaches. This paradigm emphasizes both model innovation and data equally, exemplified by the rise of foundation models [21], [22]. Despite this shift, we believe responsible AI continues to play an increasingly important role in guiding the development and deployment of AI technologies.

From a technical perspective, the ambition of responsible AI goes beyond theoretical discussions and becomes a vital aspect of both research and application [23], [24]. Therefore, this article articulates the use of AI4EO with a particular focus on responsible practices, providing a comprehensive guide for the ethical design and deployment of AI in geospatial and environmental contexts. It underscores that responsible AI is not merely a theoretical construct, but an imperative for advancing EO and geosciences. By aligning technological advances with considerations for responsible AI4EO, this discourse aims to contribute to the ongoing dialog surrounding the responsible and sustainable deployment of AI in the fields of EO and geosciences. Considering the centrality of human decisions in shaping AI applications, we emphasize the incorporation of domain-specific values at every decision point. As discussed in this article, responsible AI requires a systematic process to ensure that ethical considerations are incorporated into the body of AI4EO projects, from data collection to the interpretation of geospatial intuitions.

Figure 3 shows the main considerations of responsible AI in EO, comprising the following five important aspects:

- *Mitigating unfair bias* (*in the next section*) involves addressing and minimizing biases present in the algorithms and models used in AI4EO applications. Unfair bias can arise during many stages of the machine learning (ML) workflow, such as when the AI systems are trained on data that are not representative or when the data used for training contain inherent biases.
- *AI security in EO* (*in the section "Secure AI in EO: Focusing on Defense Mechanisms, Uncertainty Modeling, and Explainability"*) focuses on implementing measures to increase the robustness, control of uncertainty, and explainability of the AI systems and the data they process in the context of EO missions.
- *Geoprivacy and privacy-preserving issues* (*in the section "Geoprivacy and Privacy-Preserving Measures"*) aims to find a balance between leveraging the benefits of geospatial data for various EO applications while respecting and safeguarding individuals' privacy rights.
- *Ethical principles in AI4EO* (*in the section "Maintaining Scientific Excellence, Open Data, and Guiding AI Usage Based on Ethical Principles in EO"*) introduces the guidelines and standards that govern the maintenance of scientific excellence, data availability, and the guidance of AI usage in the context of EO research and applications.
- *AI4EO for Social Good* (*in the section "AI and EO for Social Good"*) refers to applying cutting-edge AI techniques to address and contribute positively to societal challenges and global issues. It aims to leverage advanced AI4EO technologies to find innovative solutions to pressing problems, promote sustainability, and enhance the overall well-being of communities.

This article presents key considerations for responsible AI4EO for the Geoscience and Remote Sensing Society. The next four sections outline the considerations necessary for implementing responsible AI methodologies within the fields of EO, the section "AI and EO for Social Good" presents the opportunities and goals relating to how a responsible AI system can be effectively utilized for social good, and the section "Responsible AI Integration in Business Innovation and Sustainability" introduces the topic of responsible AI4EO integration in business innovation and sustainability. Furthermore, we offer key remarks and actionable guidelines for researchers and practitioners to advance responsible AI practices in EO missions, as illustrated in the section "Conclusions, Remarks, and Future Directions." This review is conducted from the perspectives of both academia and industry as well as from both theoretical and practical points of view, to ensure a holistic view of the subject.

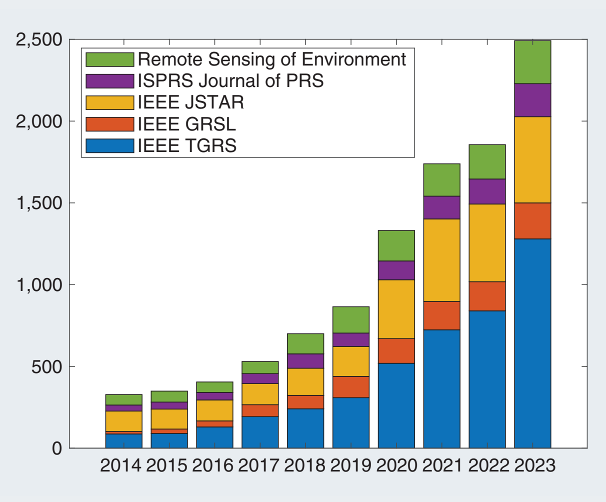

FIGURE 1. Number of AI4EO-related studies published in several GRS journals from 2014 to 2023. Data are collected from Google Scholar advanced search: keywords: ("machine learning" or "deep learning") and "remote sensing." GRS: geoscience and remote sensing; ISPRS: International Society for Photogrammetry and Remote Sensing; JSTARS: Journal of Selected Topics in Applied Earth Observations and Remote Sensing; GRSL: Geoscience and Remote Sensing Letters; TGRS: Transactions on Geoscience and Remote Sensing.

Table 1 provides the main abbreviations used in this article. This article aims to encompass a wide range of multimodal EO data, including spaceborne and airborne platforms, as well as multispectral, hyperspectral, light detection and ranging (LiDAR), synthetic aperture radar, and optical modalities. However, the primary focus is on optical data, which serve as the main source for illustrating AI4EO applications, while examples for some other modalities may not always be included.

MITIGATING UNFAIR BIAS

Biases are inherent in all ML algorithms. Indeed, courses on ML often start with the bias–variance tradeoff, where biases occur when a model is underfitting and are reduced as the model complexity increases. This reduction of model bias is paired with an increase in variance errors as the complexity of the model overfits on variances in training data that are not representative of the target class in general. "Bias" is also a term linked to the systematic errors of an algorithm or a dataset. Despite the concept of bias being inherent in ML, today there is much concern for biases present in algorithms. Indeed, a review of responsible AI guidelines and recommendations concluded that the principles of fairness and nondiscrimination should be included in all AI guidance documents

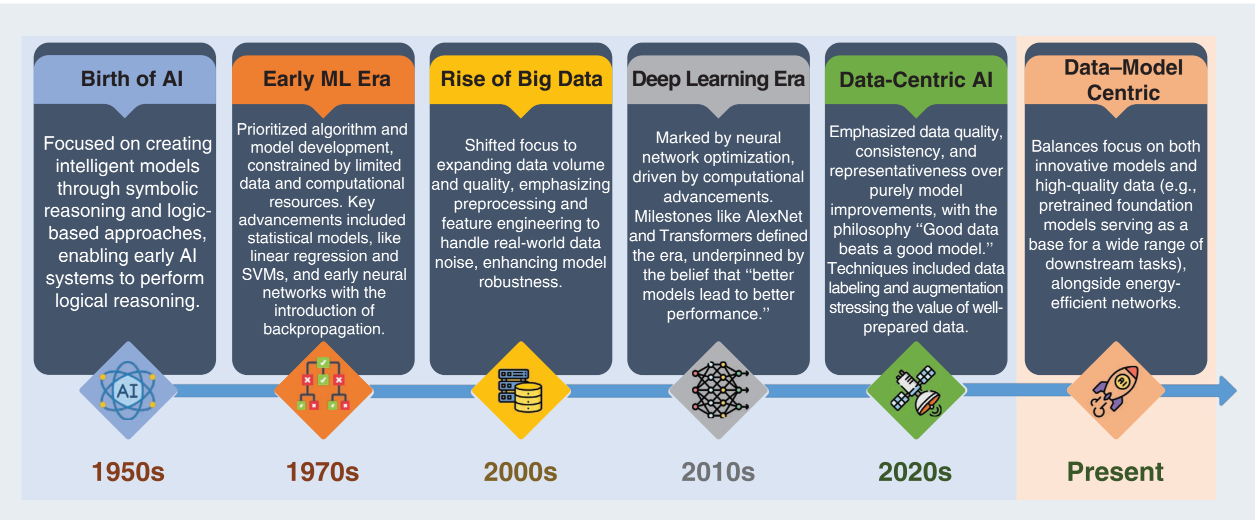

FIGURE 2. The focus of AI algorithms has shifted multiple times between being model centric and data centric. The blue part illustrates the historical development of AI concerning data and models, while the orange part reflects our beliefs in the EO community regarding current and future developments. In the 1950s, AI was introduced through early models like the Logic Theorist [3]. In the 1970s, model-centric approaches gained momentum with the development of expert systems [4] and semantic networks [5]. The 1990s saw the prominence of ML algorithms like SVMs [6], relying on datasets like MNIST [7]. The 2010s ushered in deep learning models, including CNNs and GANs, driven by large-scale datasets, such as ImageNet [8] and COCO [9]. In the 2020s, large-scale models, such as GPT [10], CLIP [11], and other transformer-based architectures emerged, underpinned by datasets like OpenAI's WebText [12] and LAION-400M [13]. Today, AI is shaped by a combination of data- and model-centric approaches, including foundation models and applications like LLMs, supported by datasets such as Visual7W [14] and OK-VQA [15]. Since the late 2010s, responsible AI has gained prominence, becoming integral to the entire pipeline—from data acquisition and model development to outcome interpretation and knowledge discovery. ML: machine learning; SVM: support vector machine; NIST: National Institute of Standards and Technology; CNN: convolutional neural network; GAN: generative adversarial network; GPT: generative pretrained transformer; CLIP: contrastive language-image pretraining; LLM: large language model; VQA: Visual Question Answering.

[25]. The difference here is that society is particularly concerned about "unfair" biases captured in ML algorithms. Generally speaking, this arises when the model systematically produces differing outcomes for specific subgroups of a class in a way that causes societal concern that is "unfair." The unfairness of the model is, therefore, contextual and closely linked to the values of a society. Friedman and Nissenbaum [26] refer to biases that cause unfair and systematic discrimination of individuals or subgroups of the population. Preventing discriminatory behavior is also considered to be the responsibility of the developer and the user of AI algorithms. The United Nations Educational, Scientific and Cultural Organization (UNESCO) "Recommendations on Ethical AI" adopted by all member states in 2021 states: "AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems" [27]. Similarly, the European Commission is drafting an AI Act that incorporates a requirement to consider and report potential biases [28]. Understanding how to audit AI workflows for biases and determining how to mitigate them are expected to be two key fields of research in the coming years.

BIASES IN ML WORKFLOWS

To understand how to identify and mitigate biases, it is important to consider how they emerge at various steps in the ML cycle [29], [30], [31]. Following the framework by Suresh and Guttag [29], six types of biases that emerge in a ML workflow are discussed (Figure 4).

1) *Historical bias* arises because field data capture phenomena that have already occurred in reality. If the state of the world changes in the future, the model may not correctly capture those changes. For example, meteorological models trained on historical data may not perform well in the future if extreme events become more prominent [32].
2) *Representation bias* is perhaps the most well-known type of bias when considering fairness. It concerns the way that the real world is sampled to generate the dataset used to train and evaluate the ML model. Representation bias occurs when certain parts of the population are underrepresented in the sampling.
3) *Measurement bias* is induced as complex phenomena are defined in simpler concepts during dataset creation. For example, when predicting poverty through EO, poverty is often defined narrowly (e.g., as a percentage of the population below a defined poverty line) rather than by

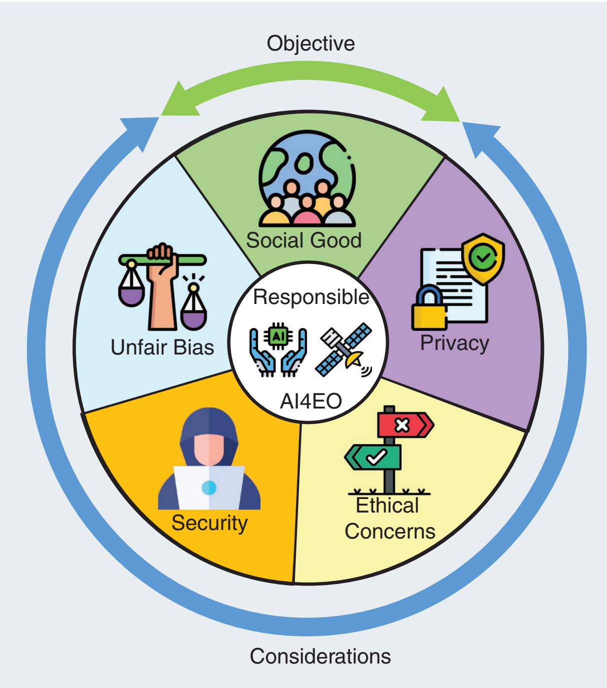

FIGURE 3. Overview of the main building blocks of responsible AI in EO. Mitigating unfair bias, securing AI (defenses, uncertainty modeling, and explainability), preserving (geo)privacy, and addressing ethical concerns outline the considerations necessary for implementing responsible AI methodologies within the fields of EO. Social good presents the opportunities and goals related to how a responsible AI system can effectively be utilized to make a positive difference in people's lives.

TABLE 1. MAIN ABBREVIATIONS.

ABBREV.	DEFINITION
AAV	Autonomous aerial vehicle
AI	Artificial intelligence
AI4EO	Artificial intelligence for Earth observation
CAM	Class activation mapping
CLIP	Contrastive language-image pretraining
CNN	Convolutional neural network
DHS	Demographic and Health Survey
DNN	Deep neural network
EO	Earth observation
ESG	Environmental, social, and governance
EWS	Early warning system
GAN	Generative adversarial network
LiDAR	Light detection and ranging
LIME	Local interpretable model-agnostic explanation
LLM	Large language model
ML	Machine learning
MoV	Mode of variability
NLP	Natural language processing
RGB	Red, green, blue
RS	Remote sensing
SDG	Sustainable development goal
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
VHR	Very high resolution
VQA	Visual Question Answering
XAI	Explainable AI

considering broader aspects, such as access to housing and sanitation [33].

4) *Aggregation bias* occurs when subgroups of a population are aggregated into a single model instead of generating different models for distinct subgroups.
5) *Evaluation bias* stems from the evaluation set used to assess model performance. Unequal representation of classes in the evaluation set as well as the use of unsuitable performance metrics can induce evaluation bias. For example, overall accuracy as an evaluation metric may be unsuitable for classification problems with unbalanced classes in remote sensing.
6) *Deployment bias* arises from the inappropriate interpretation and usage of the model results for decision making. This can occur when model results are misinterpreted or applied in contexts for which they were not intended.

AUDITING AND MITIGATION STRATEGIES

Two main streams of research can be considered in addressing biases in AI workflows: auditing for biases and implementing mitigation measures.

Auditing a workflow for biases involves identifying causality, such as determining whether a sensitive attribute causes changes in the predicted variable, or examining statistical disparities between accuracy metrics of predictions for subgroups defined by sensitive attributes [34]. For example, one might investigate whether building detection algorithms perform equally well in high-income and low-income neighborhoods of a city [35]. Calculating statistical disparities involves defining a sensitive attribute and determining which fairness metric is appropriate for the application. However, selecting a suitable fairness metric can be challenging as mathematical definitions of fairness may conflict [34], [36]. Interested readers are encouraged to explore resources on similarity-based metrics to understand the differences between metrics and how to select an appropriate one based on the application.

Bias mitigation can be applied at each step of the workflow individually. A recent review on data-centric learning in the field of geoscience and remote sensing (GRS) [37] highlights the importance of assessing dataset suitability throughout various stages of ML workflows. The review discusses how model bias can be addressed at each of these stages.

CHALLENGES AND FUTURE DIRECTIONS

There are several challenges specific to detecting biases in GRS applications.

INFERENCE OF HIDDEN SENSITIVE ATTRIBUTES IN ML MODELS

One challenge lies in defining the sensitive attributes. Societal concerns about which subgroups should be treated fairly are often connected to demographic and socioeconomic characteristics of individuals or groups. The UNESCO Recommendations [27] describe fairness according to various categories, such as age groups, cultural systems, language groups, persons with disabilities, girls and women, disadvantaged, marginalized, and vulnerable populations, rural versus urban areas, and more. Similarly, the European Union (EU) AI Act [28] is founded on Article 21 of the EU Charter of Fundamental Rights, which outlines nondiscrimination related to sex, race, color, ethnic or social origin, genetic features, language, religion or belief, minority status, age, and nationality, among others. Most of these characteristics are not included explicitly in the data utilized for GRS applications, with the exception perhaps of the urban–rural attribute mentioned by the UNESCO Recommendations. However, indirectly, these characteristics are often inferred. For example, there is extensive research on the identification of slums in satellite sensor imagery [38], [39]. Although the image does not contain information on poverty directly, the characteristics of the built-up environment have an indirect relation. Similarly, mobility data obtained through mobile phones may overrepresent higher income groups [40]. Therefore, despite not explicitly listing "sensitive" attributes in the data, geospatial and EO data are implicitly linked to these attributes. A significant challenge is to identify which sensitive attributes could be inferred from the data used to train and evaluate an ML model.

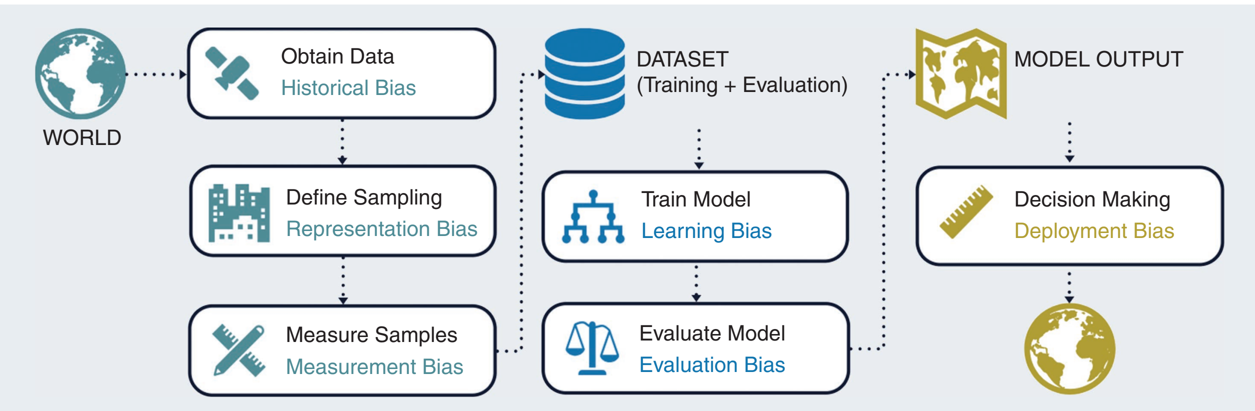

FIGURE 4. A schematic overview of the different biases that play a role during various stages of a ML workflow. (Inspired by [29].)

TABLE 2. SUMMARY OF THE THREE AI SECURITY CONCERNS INCLUDING ADVERSARIAL EXAMPLES, UNCERTAINTY, AND EXPLAINABILITY.

SECURITY CONCERN	ADVERSARIAL EXAMPLES	UNCERTAINTY	EXPLAINABILITY
Sources	Model and data-specific adversarial attacks	Error and randomness in the training and testing data, and model parameters	Complexity arising from intricate structures of the *black-box* models
Consequences	Significantly reduce the performance of trained AI models	Inability to control the reliability of model results	Untrustworthiness and nonassessability of AI models due to the invisible working process of the *black-box* models
Defense methods	Adversarial training, randomization, detection, and adversarial purification methods	Uncertainty quantification methods	XAI approaches

SPATIAL CORRELATION BIASES IN GRS

A second challenge is that of spatial autocorrelation, which is specific to the manipulation of spatial data and is often overlooked when applying methods developed in computer vision directly to geospatial data. Karasiak et al. (2022) demonstrated how the spatial dependence between training and testing samples causes an overestimation of model accuracy [41]. Other studies indicated how spatial autocorrelation plays a role in estimating the relations between air pollution and asthma [42] and in the estimation of ecosystem services [43]. However, auditing for spatial correlation biases is underrepresented in the broader literature on responsible AI in ML. It is expected that identifying and mitigating unfair biases in models will be a particularly active direction of research in the fields of GRS.

SECURE AI IN EO: FOCUSING ON DEFENSE MECHANISMS, UNCERTAINTY MODELING, AND EXPLAINABILITY

In the context of EO, ensuring AI security is vital for upholding responsible AI principles and safeguarding against potential risks and vulnerabilities. EO systems rely increasingly on AI technologies for tasks ranging from image analysis and classification to predictive modeling. However, the integration of AI introduces security concerns, such as data integrity threats, adversarial attacks on models, and privacy breaches. Addressing these challenges requires a comprehensive approach that considers both the technical aspects of AI security, such as explainable architectures and uncertainty modeling techniques, and ethical dimensions, including transparency, accountability, and the protection of sensitive information. By incorporating AI security measures into the development and deployment of AI4EO systems, we can foster trust, reliability, and societal benefit while mitigating potential harms and ensuring that AI technologies contribute positively to environmental monitoring, disaster response, and sustainable development efforts. While aspects such as ethical dimensions, mitigating unfair biases, and preserving (geo)privacy are largely explained in the sections "Mitigating Unfair Bias," "Geoprivacy and Privacy-Preserving Measures," and "Maintaining Scientific Excellence, Open Data, and Guiding AI Usage Based on Ethical Principles in EO," we focus on uncertainty modeling, adversarial defenses, and explainability in this section.

SECURITY HAZARDS IN AI ALGORITHMS

Table 2 illustrates the primary concerns regarding AI security in GRS, encompassing adversarial attacks, the uncertainty of AI predictions, and the explainability of black-box models. These significant concerns will be addressed in this section. While this article aims to emphasize concepts, we employ basic equations in this section to define AI security concerns.

ADVERSARIAL EXAMPLES

The identification of adversarial examples in AI4EO tasks [45], [46] has forced research into adversarial attacks on deep neural networks (DNNs). By introducing imperceptible perturbations to the data, these adversarial examples can deceive well-trained DNNs, leading to substantial performance degradation, even in state-of-the-art AI approaches like image classification and semantic segmentation [47] (see Figure 5). Since the deep learning-based AI4EO methods minimize loss functions as an object to train

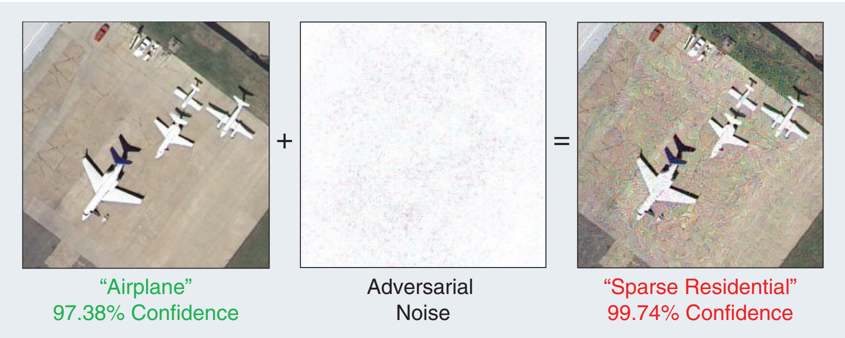

FIGURE 5. Illustration of adversarial attack: a minor perturbation can fool the classifier into wrong predictions [44].

the models, adversarial attacks employ the idea of maximizing the loss functions to find a perturbation to contaminate the clean samples, which can consequently fool well-trained AI4EO algorithms. As a result, this behavior severely threatens the responsibility of the current AI4EO approaches by making them predict incorrect results with high confidence.

Adversarial attacks can appear virtually and physically when applying deep learning models. On the one hand, the attackers can directly modify the applied remote sensing data before they are forwarded to the application models. For example, Wu et al. [48] demonstrated that the current end-to-end autonomous driving system could be easily broken by tempering the data with mild perturbations generated by white-box attacks. On the other hand, physical attack methods are more imperceptible since the adversarial perturbations are introduced during the remote sensing imaging process [49]. Zhang et al. [50] explored physical attacks in real-world AI4EO applications by applying an adversarial patch attack on multiscale object detection for autonomous aerial vehicle (AAV) remote sensing images. As a corollary, responsible AI is required to mitigate the potential threats of these adversarial perturbations. Adversarial defense methods are desired that improve the reliability and response of the model; they are introduced in a later paragraph in this section.

THE INHERENT DATA UNCERTAINTY DUE TO DOMAIN VARIANCE REMAINS A CHALLENGE THAT CANNOT BE COMPLETELY ELIMINATED.

UNCERTAINTY OF THE MODEL PREDICTIONS

GRS data, often sourced from varied platforms, exhibit diverse domain distributions, introducing uncertainty in AI model predictions, as described in the section "Mitigating Unfair Bias." The stochastic nature of ML models further amplifies this uncertainty, posing a significant challenge to prediction reliability [51]. Recognizing and managing uncertainty is vital for the development of responsible AI4EO models. The application of AI in GRS data involves distinct training and testing phases. To this end, uncertainties may arise in the samples of training and testing datasets as well as in the model parameters, categorized into data and model uncertainty [52], [53].

Lack of domain knowledge is a prominent factor contributing to data uncertainty, arising from the disparate domain distribution observed between the training dataset and the testing dataset. In remote sensing imaging, spectral features of observed data are closely tied to spatial and temporal conditions, such as solar illumination, season, and weather. Consequently, variations in these conditions can markedly influence imaging outcomes, resulting in heterogeneous data with diverse domain distributions, a phenomenon known as *domain invariance* [54]. Regrettably, existing AI approaches struggle to make accurate predictions with a decision boundary that accounts for data exhibiting domain invariance, often referred to as *domain shift* [55]. This poses a significant threat to model performance, particularly when faced with data uncertainty in testing samples during the model deployment stage.

In attempts to mitigate domain shift issues, numerous studies in GRS have proposed methods to adapt and extend the decision boundary of classification models to encompass unknown data in the testing phase [56], [57]. In particular, metalearning has been employed to fine-tune pretrained models, enabling them to generalize across diverse, previously unseen data distributions by minimizing domain discrepancy [58], [59]. Additionally, incremental learning approaches have been introduced to help models adapt to evolving categories in new, out-of-distribution data [60], [61]. However, these methods are limited in their ability to perfectly fine-tune models, and the inherent data uncertainty due to domain variance remains a challenge that cannot be completely eliminated.

In addition to data uncertainty, model uncertainty encompasses the errors and randomness inherent in the model parameters, initialized and optimized during the model design and training processes. As researchers strive to develop advanced AI algorithms with cutting-edge performance and efficient training, diverse model architectures linked with optimization strategies have been devised for various GRS applications [62]. However, determining the optimal model architecture and training hyperparameters remains challenging, introducing uncertainty into the trained model that can subsequently impact predictions, commonly referred to as *model uncertainty* [63].

OPACITY OF BLACK-BOX MODELS

As high-performance computing technology advances, contemporary AI applications in GRS studies explore models with deeper architectures and more complex parameters to enhance classification performance. However, the increased complexity raises challenges in understanding and interpreting the internal operational steps of these architectures, rendering the models akin to black boxes [64]. Compared with traditional ML approaches, the opacity of these black-box models hinders the identification of hidden security hazards that can significantly impact classification as well as the reasons for "wrong" predictions in real-world applications [65]. Moreover, the growing emphasis on data privacy and high confidentiality in many GRS tasks underscores the need for trustworthy AI models aligned with ethical and judicial requirements for both developers and users.

ROBUST MODELING FOR AI SECURITY

In response to the identified security hazards outlined in the preceding section, researchers are working toward the development of more robust AI models, emphasizing enhanced responsibility for AI security. Several strategies are being employed, including adversarial defense, uncertainty quantification, and explainable AI (XAI), to address challenges related to adversarial examples, data and model uncertainty, and black-box models, respectively.

ADVERSARIAL DEFENSE

Common adversarial defense approaches include adversarial training, randomization, detection, and adversarial purification methods. Adversarial training methods aim to bolster the model's robustness against adversarial examples by incorporating them directly into the model training stage alongside normal training samples. For example, Goodfellow [66] proposed an empirical adversarial training scheme involving the inclusion of fast gradient sign method. However, it is important to note that while adversarial training methods may enhance the model's robustness against specific attack methods, vulnerabilities against adversarial examples of other types may persist.

Given that adversarial perturbations can be perceived as additive noise on adversarial examples, randomness methods introduce random components to enhance the model's robustness. For example, Sitawarin et al. developed a randomized smoothing technique for adversarially robust classification involving the randomization of the input to the DNN to eliminate potential adversarial perturbations in the input samples [67]. However, it is important to note that the performance of randomness methods depends greatly on the quality of the randomization process and can be affected significantly by both the lack of prior knowledge of attack types and theoretical explanations.

UNCERTAINTY QUANTIFICATION

To better comprehend and manage potential uncertainty, uncertainty quantification techniques have been proposed to assess the credibility of predictions. In many GRS applications, classification results manifest as distributions of possibilities associated with object classes. These distributions are typically transformed by a softmax function at the final stage of the network. Grounded in the principles of Bayesian inference, the data uncertainty of an input sample can be described as a posterior distribution over the class label, given a set of model parameters. Similarly, the model uncertainty can be formalized as a posterior distribution, given the training dataset.

On this basis, various uncertainty quantification approaches aim to obtain an uncertainty distribution of model predictions by marginalizing the model parameters. These methods fall into two broad categories: deterministic methods and Bayesian inference methods, differing in subcomponent model structures and error characteristics. In deterministic methods, the parameters of AI models are deterministic and remain fixed during the inference step, aligning with the common behavior of most developed AI models. By contrast, Bayesian inference methods employ the Bayesian learning strategy, leveraging the ability to seamlessly integrate the scalability, expressiveness, and predictive performance of neural networks for uncertainty quantification [68], [69].

XAI

In general, XAI aims to provide an understanding of the pathways through which output decisions are made based on the parameters and activations of the trained models. To this end, many XAI approaches have been proposed for a comprehensive understanding of the model behavior to justify the individual feature attributions of a test sample, which provides a powerful tool to examine incorrect predictions and inspect the potential threats of black-box models, especially in high-risk GRS applications [70]. For example, Ribeiro et al. [71] developed a local interpretable model-agnostic explanation (LIME) method to provide understandable representations for the predictions of black-box classifiers by highlighting attentive contiguous superpixels of the source image with positive weight toward a class, which benefits users in learning "how the model thinks" when classifying a specific image. Furthermore, the deep features of DNNs can also be processed into visualizations to observe how the training procedure works on specific tasks, as shown in Figure 6.

CHALLENGES AND FUTURE PERSPECTIVES

In this section, we discuss the future perspectives of AI security for EO tasks from three perspectives.

THE GAME BETWEEN ADVERSARIAL ATTACK AND DEFENSE TOWARD EO APPLICATIONS

While adversarial examples have been identified widely in several EO applications, research on adversarial attacks and defense algorithms has been developed and raised debate in game theory. On the one hand, the attackers have proposed more powerful attack algorithms to affect a wider range of DNNs applied in various EO missions, including scene classification and semantic segmentation. On the other hand, the defense group has also identified the emerging adversarial threats and tended to secure AI algorithms from more complex adversarial examples with frontier techniques, such as diffusion models [44]. In conclusion, the game between adversarial attack and defense algorithms has been started and extended to more subtopics in the AI4EO research field [73].

PRACTICAL UNCERTAINTY ANALYSIS FOR AI4EO

The uncertainty in both model and data toward AI4EO missions has been a critical obstacle for developing AI models with more reliability and responsibility. Although some uncertainty quantification methods have been seen in GRS studies, most of them were designed for dedicated models and are not adaptable to most of the existing AI4EO approaches. To this end, universal uncertainty analysis with greater capability to be applied in more typical DNNs should be developed for AI4EO missions.

XAI SCHEMES WITH MORE UNDERSTANDABLE INTERPRETATIONS

Explainability is considered to be a significant attribute of modern AI4EO missions providing the opportunity for humans to learn the intrinsic process of the AI algorithms, so that security risks can be largely alleviated. However, current XAI methods designed for EO missions can only generate visualization interpretations, which are usually semantically misaligned and still too abstract to understand directly. Therefore, future XAI studies should focus on increasing the depth and accuracy of XAI results by introducing more constraints and optimization problems to explanations.

GEOPRIVACY AND PRIVACY-PRESERVING MEASURES

Remotely sensed images and geospatial data offer unprecedented opportunities to monitor our planet and support spatial planning, management, and decision making. Nevertheless, as remote sensing technologies evolve, the need to balance the advantages of data-driven decision making with the protection of individual privacy becomes paramount. The increasing resolution and quality of remotely sensed data, with their ability to capture detailed information about Earth's surface and (directly or indirectly) the people living in those areas, raise concerns about the potential misuse of data and the accidental revelation of sensitive information [74]. AI technologies allow the scale-up of geospatial applications, reducing analysis costs and subjectivity and, at the same time, magnifying the risk of exposing sensitive personal information.

Effectively addressing these challenges requires thoughtful consideration of ethical principles, legal frameworks, and technological safeguards to ensure that the benefits of geospatial advances are realized without compromising individual privacy rights. Earlier research endeavors aimed at constructing a conceptual privacy framework; for example, [75] identified seven different concepts of privacy by examining distinct emerging technologies.

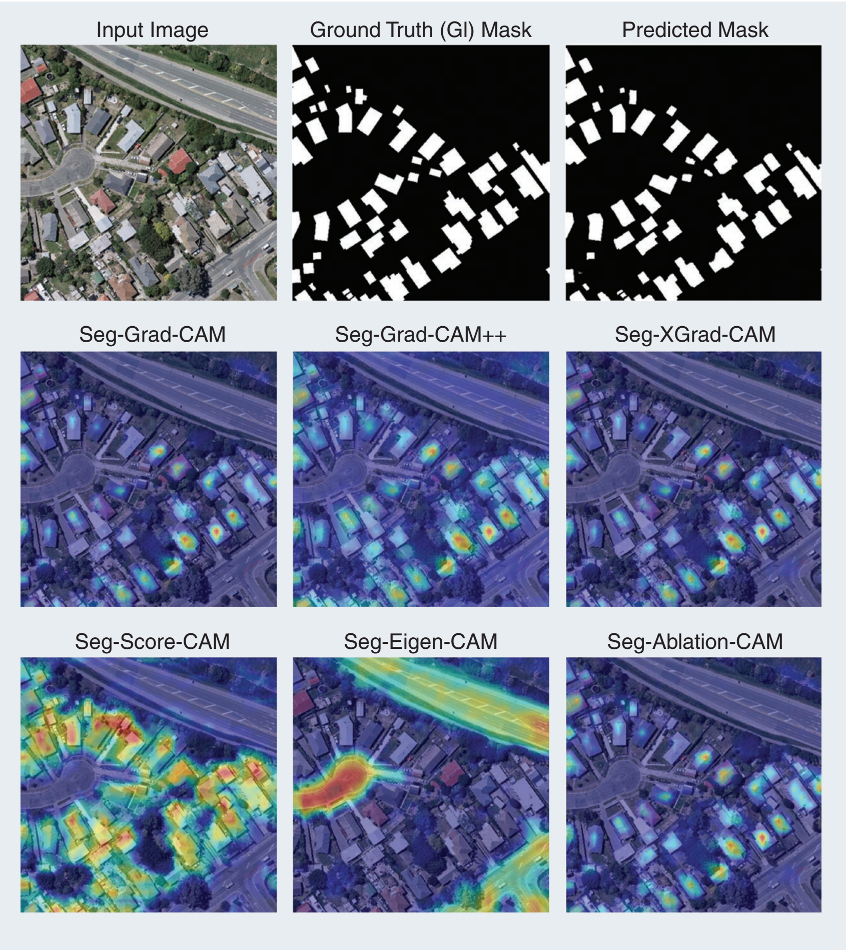

FIGURE 6. Heatmaps of UNet decoder block with different CAM-based XAI algorithms for the class of "buildings" in the semantic segmentation of a remote sensing image. Pixels that have a brighter color indicate a higher probability of being classified into the target class. [Source: [72].]

PRIVACY IN AAV IMAGES

In remote sensing, privacy concerns often arise when the resolution of an image is sufficiently fine to discern individuals' identities and expose personal details. This scenario commonly occurs during the analysis of images captured by AAVs, potentially encroaching upon 1) location privacy as individuals can be identified and tracked within AAV images; 2) behavior privacy within private settings, free from external surveillance; 3) space privacy, revealing details about secluded areas like backyards; 4) association privacy, disclosing group memberships and affiliations; and 5) data and image privacy, involving individuals' control over images featuring their presence.

Gevaert et al. [76] evaluated the societal impacts of using AAVs for informal settlement mapping through two case studies in Eastern Africa. The study stresses the importance of notifying citizens of the purpose of data collection, the rights to access, and data processing and distribution, especially when flying over marginalized communities. The absence of a cohesive policy framework governing such activities shifts a significant portion of responsibility onto industry self-regulation, potentially falling short of adequately safeguarding marginalized communities. Nevertheless, the existence of a "unified" framework is challenged by the specific application of AAV image acquisition and the fact that the concept of sensitive information or privacy may vary among people, groups, and cultures. The study illustrates the importance of considering the local context regarding privacy issues. Another example application with potential privacy-related concerns is the use of AAV images in support of first responders in emergency situations (e.g., after an earthquake or other types of disasters). The authors of [77] developed a deep learning network to detect victims in AAV images, which is trained to detect humans even in complex situations where body parts of the victims are partially covered by dust or debris. However, collecting a dataset for training such a network is challenging because of the sensitive nature of the data. Therefore, the authors proposed a framework to generate a harmonious composite of images for training. The framework first pastes images of body parts onto a debris background to generate composite victim images and then uses a deep harmonization network to make the composite images more harmonious.

MAPPING SENSITIVE INFORMATION

Besides decimeter/centimeter-resolution AAV images, privacy concerns may also arise with satellite sensor images of various resolutions, especially when applied to map societal and socioeconomic features. Examples of sensitive applications include the mapping of informal settlements, slums, deprived urban areas [78], [79], [80] and refugee camps [81]. The peril associated with disclosing sensitive information is evident in the "Mass Atrocity Remote Sensing" initiative, which sought to furnish forensic evidence of crimes against humanity through the analysis of satellite sensor imagery and other datasets. Nevertheless, researchers soon recognized that these data could also be exploited to identify and harm individuals and groups. Publishing sensitive maps should always be done with care to prevent unintended consequences for local communities. In the case of informal settlements, being classified as living in such a neighborhood may result in social exclusion or even exposure to eviction (e.g., see the reports of forced evictions of 30,000 people from Nairobi's largest slum, Kibera, in 2018 for building a new road [82]). Measures for privacy preservation typically include downgrading map spatial resolution (e.g., to cells of 100 × 100 m), aggregating data at the administrative unit level, and removing geolocation from datasets.

Another strategy to preserve household privacy is to apply jitter to geospatial point data (adding random noise to geocoordinates). The authors of [83] developed a comprehensive and openly accessible collection of microestimates detailing the distribution of relative poverty and wealth across 135 low- and middle-income countries. They gathered data from Demographic and Health Surveys (DHSs) through traditional face-to-face surveys conducted with 1,457,315 distinct households residing in 66,819 villages spread across 56 different low- and middle-income countries globally. The DHS data contain approximate GPS coordinates indicating the centroid of each surveyed village among the 66,819. However, to safeguard privacy, the precise geocoordinates are obfuscated, with a margin of error of up to 2 km in urban areas and up to 5 km in rural areas.

CHALLENGES AND FUTURE PERSPECTIVES

A recent challenge is the increasing availability of satellite sensor images with submeter resolution up to 30 cm (e.g., WorldView-3, Pleiades NEO, and SkySat) able to capture potentially sensitive information. The potential of such data for mapping and object detection should, therefore, be counterbalanced with privacy-preserving measures. Moreover, it is worth noting that privacy requirements may clash with open science principles, restricting the possibility of openly publishing and benchmarking datasets.

ANOTHER STRATEGY TO PRESERVE HOUSEHOLD PRIVACY IS TO APPLY JITTER TO GEOSPATIAL POINT DATA (ADDING RANDOM NOISE TO GEOCOORDINATES).

Future investigations should develop strategies to openly share data for scientific advancements while preserving sensitive personal or group information. Such strategies are much needed in applications such as the mapping of deprived urban areas or victim detection, where further advances in the development of AI-based methods critically depend on the availability of large benchmark datasets, which are currently missing.

MAINTAINING SCIENTIFIC EXCELLENCE, OPEN DATA, AND GUIDING AI USAGE BASED ON ETHICAL PRINCIPLES IN EO

The real-world character of AI4EO applications with downstream political, societal, and environmental impacts offers tremendous opportunities for positive change while simultaneously bearing the risk of negative impacts when not carefully designed and implemented [84], [85], [86]. Maintaining the highest scientific standards and making data and code openly accessible while complying with ethical principles are essential prerequisites to enhance the transparency of knowledge production, ensuring the high reproducibility of scientific work and maximizing the societal, economic, or environmental benefits of AI4EO research, development, and applications [84]. Compliance with fundamental ethical principles is, thus, a major duty of the scientific community.

A series of review articles laid out the relevant definitions and concepts of ethics in AI. For example, Jobin et al. (2019) [87] identified five broad categories from ethical guidelines in AI research: transparency, justice and fairness, nonmaleficence, responsibility, and privacy. With a focus on AI4EO, Kochupillai et al. (2022) [88] laid out six ethical dimensions: privacy, honesty, integrity, fairness, responsibility, and sustainability (for an overview on recent AI ethics categorizations, see Table 2 in [88]). We here revisit the six categories proposed in [88], focusing on aspects that transcend the themes of maintaining scientific excellence, open data, and guiding AI usage in AI4EO. These include those that directly or indirectly relate to workflows, defining objectives and research questions, data inputs and analysis outputs, and their downstream impacts.

PRIVACY

In the context of responsible AI4EO, we situate privacy concerns arising under the open data paradigm and with regard to (territorial) stigmatization. Open access to datasets and code is a core element of scientific reproducibility. Scientific excellence in AI4EO can, thus, only be achieved when the public dissemination of essential research data and code is made available to enable full reproducibility and robustness checks [89]. The scientific community has pushed for improving the *findability*, *accessibility*, *interoperability*, and *reuse* of digital assets, which materialized in the FAIR principles [90]. Increasingly, however, the publication of geospatial data, including AAV or fine-to-very-fine-resolution satellite sensor imagery, labeled datasets, or the outputs of AI4EO workflows, may raise privacy concerns [91]. The FAIR principles demand data to be free from use restrictions, fees, or embargos, but legal, moral, or ethical reasons, such as safeguarding individual privacy or endangered species, may justify deviation from these best practices on data accessibility [92]. However, detailed metadata should be made available enabling the discovery of such data and describe explicitly the conditions under which data access can be granted.

WE HIGHLIGHT THAT COOPERATION WITH REGIONALLY KNOWLEDGEABLE PARTNERS IN THE EARLY PROJECT DESIGN PHASE IS AN EFFICIENT WAY TO AVOID THE PITFALLS OF STIGMATIZATION IN RESPONSIBLE AI4EO.

Tradeoffs between the benefits of open data and privacy concerns therefore need to be evaluated case by case to avoid privacy-violating or other ethically concerning impacts of public data dissemination. Besides the techniques to obscure geolocation information described previously, a selective disclosure of data upon requests may be justified to avoid risks of data misuse. Additionally, purpose-specific data licensing may provide opportunities to avoid the unintended use of datasets. One example of purpose-oriented data releases is Norway's International Climate and Forests Initiative Satellite Data Program, which provides commercial satellite imagery covering the world's tropics for noncommercial sustainability-centered applications (https://www.planet.com/nicfi/).

Stigmatization may occur in cases where urban neighborhoods are categorized as slums [93], [94], or categorizations of crime rates or poverty levels adversely affect public perceptions of a community [95]. The resulting territorial stigmatization may influence decision making and negatively affect the local population. Similarly, stigmatization applies to AI4EO applications that fail to consider the realities on the ground, for example, those designed to expose illegal activities, such as deforestation, fire use, or mining activities, without knowledge about potential concessions to conduct the respective activity. Similarly, stigmatization may occur when global or broad-scale analyses fail to sufficiently consider local realities. For example, small-scale farming has been found to be responsible for 97% of deforestation on the African continent [96]. However, regional dynamics, such as the rapid emergence of medium-scale commercial farming operations [97], [98], the increasing role of export-oriented commodity crops, such as cocoa, cashew, and oil palm in deforestation [99], and tradeoffs among environmental degradation, food production, livelihoods, and labor opportunities in semisubsistence agriculture [100], [101] allow for a more nuanced perspective of such dynamics. As such, avoiding stigmatization of people, communities, or localities requires a high level of awareness of researchers in the design of research questions and label heuristics as well as embedding scientific findings and their limitations in the realities of the region under investigation [102]. We highlight that cooperation with regionally knowledgeable partners in the early project design phase is an efficient way to avoid the pitfalls of stigmatization in responsible AI4EO [103].

HONESTY

Truthfulness and trustworthiness are inherent to scientific inquiry, but the current replication crisis undermines the credibility of science and, thus, hampers the societal uptake of scientific outputs. As a community producing knowledge with real-world impact, responsible AI4EO requires transparency, explainability, and data veracity [84].

In AI4EO, transparency not only requires thorough documentation and referencing of the datasets and workflows used but also the need to understand and clearly communicate ambiguities, potential biases and errors, or conceptual limitations. In light of the ongoing democratization of EO data and workflows through cloud platforms and accessible user interfaces, empowerment through access to EO data requires a thorough understanding of the limitations of EO data and workflows. Furthermore, transparency enables critical reflection of scientific findings. One example is a recent study linking the observed emerging abilities of large language models (LLMs) (sudden, unexpected increases in performance) to study design choices made by researchers (e.g., evaluation metrics) [104].

Assuring explainability in AI4EO continues to be challenging, but it is a prerequisite for early diagnostics regarding unwanted bias and sources of error. We strongly encourage the community to conduct research on XAI and incorporate it, wherever possible. Focusing research on explainable AI4EO will contribute to the recognition of EO data as a distinct modality in the AI community at large by providing a better understanding of how AI models leverage the information contained in EO data [85]. Greater explainability may also profit the emerging subfield of self-supervised learning [105] and help to ensure oversight of (geographic) biases in AI4EO foundation models that claim high degrees of geographic generalization but are trained in selected world regions [106]. Such biases are, for example, present in LLMs [107] and may lead to error propagation to many downstream applications [108].

Data veracity in EO concerns both the imagery and the EO data preprocessing workflows used as well as reference data/labels being in accordance with a (nonstigmatizing) heuristic defining the objects of interest. Using well-established and tested processing pipelines to assure the highest possible quality of EO data, for example, in terms of atmospheric correction [109], but also regarding quality assurance flags, including cloud masks for optical data [110], is recommended to avoid downstream errors related to poor data quality. Labeling (the process of categorizing real-world phenomena into distinct classes or groups) is highly context specific. Therefore, it requires regional expertise and knowledge, is ideally coordinated with local experts, and can be considered an iterative process in which exchange of knowledge from various disciplinary lenses is an effective means to reduce the risk of false categorizations or stigmatization [102].

INTEGRITY

Technical robustness and uncertainty reporting are key elements of integrity in AI4EO. Detailed documentation and reporting of the employed processing pipelines, including the respective processor version, will enhance the reproducibility of AI4EO analyses. Detailed reporting of biases, errors, and uncertainty adds credibility and is helpful in avoiding erroneous use of datasets in downstream applications. Given the increasingly close interactions between the ML and EO communities, the maintenance and use of well-established evaluation standards particular to the EO domain, such as in reporting map accuracies and area estimates in categorical settings [111] or for detecting trends in spatiotemporal data [112], is key.

In AI4EO research, a focus on single aggregate evaluation metrics (e.g., overall accuracy), as discussed in the section "Mitigating Unfair Bias," is insufficient as often the detection of minority classes is most relevant in many AI4EO applications. Generally, and for imbalanced problems in particular, reporting area-adjusted class-specific user and producer accuracies from an independent and randomized sample must be preferred over simplistic 1D statistics. Similarly, class area estimates are to be derived from reference samples rather than pixel counting in maps with nonrandom error distribution. For continuous targets (e.g., reflectance, biomass, crop yields, or canopy height), reporting multiple error metrics and assessing them across the entire value range reveals heteroscedasticity and can inform downstream users of potential pitfalls when using a product [113].

FAIRNESS

Fairness here refers to avoiding bias and to respecting all aspects of diversity and sociocultural differences, ultimately creating standards based on principles related to fairness. Bias may relate to all stages of AI4EO analyses, beginning with defining a research question, through the workflow itself, and ending with the manifold downstream impacts related to interpreting the results—as described in the section "Mitigating Unfair Bias." We will not delve deeper into the broader ethical framework discussed elsewhere [88], [114]. Instead, we focus on core needs before, during, and after setting up an AI4EO workflow.

ASSURING EXPLAINABILITY IN AI4EO CONTINUES TO BE CHALLENGING, BUT IT IS A PREREQUISITE FOR EARLY DIAGNOSTICS REGARDING UNWANTED BIAS AND SOURCES OF ERROR.

Before setting up an AI4EO workflow, we may consider opportunities to invite potential stakeholders of anticipated outcomes to participate in cocreating the research agenda [115]. Stakeholders typically include different institutions across all levels, from village heads in an area of interest to the United Nations or the World Bank. Stakeholders may also be citizens who are acquainted with the topic or the area of interest, ideally (but not mandatory) being engaged in related citizen-science opportunities [116], such as data sampling or field-based mapping evaluation [117], [118]. Citizens and institutions acquainted with the research questions will help to mitigate bias in the project design phase.

During workflow creation, interpretable AI, as discussed in the section "Secure AI in EO: Focusing on Defense Mechanisms, Uncertainty Modeling, and Explainability," should support paving the way toward optimizing workflows and avoiding downstream misinterpretations [119], [120]. Interpretable AI also allows tuning algorithms and workflows based on the expertise of both scientists and nonscience stakeholders [121]. Nonscientists thereby provide consultancy to EO scientists to overcome the different biases mentioned in the section "Mitigating Unfair Bias" during the workflow implementation and rollout, including historical, representation, measurement, aggregation, and evaluation bias (compare with the section "Biases in ML Workflows"). Next to training data creation, providing sufficient and independent evaluation data for accuracy assessment is both of core importance and often one of the most challenging tasks (see also the section "Integrity").

Implicitly, avoiding workflow-related bias is also tightly connected to recognizing deployment bias. Deployment bias relates to the inappropriate interpretation of results after the data analysis, which is often related to downstream decision makers. This underlines the importance of cocreation and coproduction to minimize and mitigate deployment bias early on. In other words, confronting stakeholders with results created in the scientific "ivory tower" without engaging them in the design and production phase regularly creates unsatisfactory results for those who are affected or are supposed to build their decision-making processes on AI4EO outcomes.

Deployment bias also points directly to the need to respect diversity and sociocultural differences. For example, we need to recognize different ethnic or religious communities as well as AI4EO experts compared to nonexpert stakeholders to avoid misinterpretations or disrespectful outcomes of AI4EO workflows. Typical examples are the mapping of slums or agricultural practices in the Global South [93], [122]. Many studies on related research questions need to integrate stakeholders early on in AI4EO to include the necessary expertise to create meaningful results, specifically across large areas of interest, where diverse and multicultural stakeholders are needed. AI4EO workflows also rarely consider resource restrictions in the Global South—regarding the "data desert," limited processing resources, or limited financial backup, for example, for implementing cloud computing or using high-performance computing or supercomputing facilities [123]. In other words, cooperation at "eye level" is needed to embed AI4EO in the proper sociocultural context, specifically when solving sustainable development goals (SDG)-related questions and global change issues in the Global South [103], [119]. In conclusion, we create win–win situations when joining the forces of stakeholders "on the ground" (deliberately including stakeholders across scales, from local dwellers to global institutions, depending on the research questions to be tackled) with AI4EO experts.

THERE IS INDEED A LACK OF UPSCALING STUDIES DEMONSTRATING THE SUITABILITY OF MOST ALGORITHMS FOR TACKLING REAL-WORLD PROBLEMS.

RESPONSIBILITY

The need for interdisciplinary and regional collaboration related to domain knowledge across the entire workflow of AI4EO and unwanted outcomes or interpretations was tackled in the section "Fairness." Therefore, here we focus on effective handling of the distinctness of EO data [85].

It has been argued that the AI/ML community urgently needs to recognize the distinct features of satellite-based EO data and to "move out of the comfort zone" of applying AI methods established in other domains without improved adaptation to EO or satellite data ML [85]. The reasons are manifold. For example:

- Satellite data are rarely "analysis-ready data" [124]; they require state-of-the-art preprocessing to represent the same values for the same features in different images, i.e., to avoid pseudovariance.
- Spatial autocorrelation according to Tobler's first law of geography [125] is rarely used to improve AI4EO, and even less so newer concepts like partitioned autoregressive time series that consider both spatial and temporal autocorrelation in a congruent fashion [112].
- Satellite data provide a wide range of spatial resolutions (from very-fine resolution to pixels of several square kilometers) and extend across all scales to be observed (from local, fine-scale objects in urban environments to global phenomena). Both very diverse spatial detail and multiscale research ask for adapted solutions.
- Different temporal resolutions, i.e., data densities over time, represent the same land surface differently. In other words, change analyses over years or decades are far from trivial and need contextualization [113], [126]. This is even more important when phenology is considered, and specific temporal windows across the seasons are crucial to identify the relevant phenomena, for example, different crop types, land-use intensities, or tree species separable only at certain phenological stages [127], [128], [129].
- Most AI algorithms for image analyses were developed on three-band red, green, blue (RGB) imagery, often based on very high resolution (VHR) drone data. Today's multispectral-to-hyperspectral satellite data ask for more sophisticated approaches to detect not just the obvious, but the relevant. This is aggravated by more spectral detail in radiometric data beyond 8-bit resolution and even more by using virtual constellations [130], ranging from optical to radar to LiDAR data.
- Sparse labeling problems occur with increasing areas of interest enabled by AI4EO. More severe problems are inherited by biased samples across larger regions or nontrivial features to be sampled from existing data or the imagery itself [131], [132].

While all of these aspects are relevant for any EO analysis framework, opportunities have increased rapidly with AI4EO (e.g., regarding the size of areas of interest that can be handled). However, along with these come related risks and potential awareness deficits. Many, if not the majority of, AI4EO studies focused on comparably small use cases in the past to develop or apply algorithms. However, the rollout to real-world problems, for example, at the national, continental, or even global scale, requires ubiquitous algorithms that can handle much more diverse feature spaces (over time). There is indeed a lack of upscaling studies demonstrating the suitability of most algorithms for tackling real-world problems [85], [133], [134].

SUSTAINABILITY

Since 2015, the UN has set an ambitious agenda related to the SDG, aiming to improve social, economic, and environmental conditions globally by 2030 [135], [136], with a wealth of implications for AI and also specifically for AI4EO [123]. Here, we take the broadest possible standpoint, including the intrinsic sustainability of AI4EO workflows and the external impacts for all three SDG dimensions. Given that AI at large offers the potential for both positive and negative impacts across all SDGs, with a majority on social sustainability [137], AI4EO has a similar potential for both positive and negative impacts.

Focusing on environmental sustainability, EO is both data and energy hungry, thereby creating immense CO_2 emissions. This footprint relates to the entire lifecycle of EO engineering, science, and outcomes, including, but not limited to, building, launching, and running satellite missions, data creation and curation, data storage, and data analysis workflows—the data-related aspects being core for AI4EO [138].

Zooming in on economic and social SDG, the FAIR principles are highly relevant: we have already thematized avoiding "helicopter research" in the section "Fairness" [103]. Offering inclusiveness is nontrivial to deliver as many EO scientists and companies, specifically in the Global South, do not have access to powerful cloud computing or high-performance computing services or to GPU-equipped high-performance workstations. Even if access to online services may become available through dedicated programs or international cooperation, proper education and training for the next generation of AI4EO scientists is of utmost importance [123].

The outcomes from AI4EO research may greatly support improvements along the lines of the UN SDGs [136], [137], but they may also lead to ill-posed conclusions. Such conclusions can, in the worst case, create an existential economic threat (e.g., if loans or insurance are not provided based on AI4EO-derived indicators [139]). Also, local stakeholders, including indigenous people, who may not even be aware of government decision making processes, are regularly impacted. Last but not least, AI4EO may fall into the "wrong hands"; for example, it may be used by autocrats, dictators, or criminals to suppress or threaten specifically environmental and social SDGs. The most disturbing examples of this are the occurrences of ubiquitous AI4EO-supported surveillance and warfare.

AI AND EO FOR SOCIAL GOOD

In this section, we build upon the discussions in the section "Maintaining Scientific Excellence, Open Data, and Guiding AI Usage Based on Ethical Principles in EO" regarding the use of AI4EO for social good, which involves applying AI technologies to address and solve social, economic, and environmental challenges, such as health care, education, environmental sustainability, poverty alleviation, and disaster response. By leveraging AI technologies and the ethical considerations mentioned in the previous sections, these initiatives seek to improve outcomes and make a positive difference in people's lives, which is the ultimate goal of responsible AI.

Although the applications of AI4EO for social good are extensive [136], [141], we focus here primarily on understanding disaster risks and evaluating their impacts through several key examples crucial for effective disaster management and financing strategies [142], [143], [144]. Robust frameworks are necessary for investing in projects that enhance resilience to climate change and its associated disasters. AI, leveraging integrated EO data, can significantly contribute to building these frameworks by analyzing such data through innovative computational methods [142], [143], [145]. However, it is also crucial to implement policies and decisions at the local level as the impacts of climate change vary over time and across different regions [142], [146], [147], [148], [149].

Understanding disaster risks and evaluating their impacts are crucial for effective disaster management and financing strategies [142], [143], [144]. Robust frameworks are necessary for investing in projects that enhance resilience to climate change and its associated disasters. AI leveraging integrated EO data can contribute significantly to building these frameworks by analyzing such data through innovative computational methods [142], [143], [145]. Moreover, it is also crucial to address the implementation of policies and decisions at the local level as the impacts of climate change vary over time and across space [142], [146], [147], [148], [149]. For example, rising temperatures in Europe can intensify evapotranspiration, potentially causing more floods in Germany and water shortages in Italy. Portugal, France, and Greece may face increased wildfires caused by similar processes [142], [146], [147], [150]. Global warming also impacts the water cycle, necessitating increased irrigation for vegetation health and food security. This can heighten the occurrence of natural hazards, even across significant distances. For example, irrigation in North America's Central Valley, High Plains, and Mississippi River Valley affects precipitation in regions like the Colorado River, the U.S. Midwest, and the U.S. Southeast, respectively [151], [152], [153], [154].

The preceding instances emphasize the diverse impacts of climate change on different regions, even at relatively close range. Thus, it is crucial to enable each region to be prepared and ready to implement specific contingency plans based on local changes to environmental events characterized by changes in atmospheric and oceanographic variables [142], [144], [146], [147]. It is also worth noting that these changes might affect the socioeconomic fabric of diverse biogeographic regions, thereby influencing the stability and security of human welfare and demographic fluxes [142], [143], [144], [146], [147], [148]. It is, therefore, paramount for AI to provide a deep understanding of EO data so that decision makers and policy makers are enabled to implement climate change adaptation and mitigation strategies. These strategies affect investments and solutions that citizens, local authorities, and businesses can undertake. Additionally, it implies the ability to engage stakeholders and public opinion to gain societal consensus [145], [146], [147], [150].

To provide concrete examples of the actual impact and possibilities unlocked by the appropriate use of AI in decision-making processes, we present two instances of operational scenarios that would benefit greatly from a deep understanding of Earth's processes to enhance human welfare and improve communities' well-being.

EARLY WARNING SYSTEMS FOR MASS MOVEMENTS

One of the most severe effects of climate change is the increased occurrence of catastrophic events caused by mass movements. This affects human welfare and results in huge economic losses. As such, the development of early warning systems (EWSs) is crucial to the everyday decisions of local governments and communities and for exemplifying responsible AI by prioritizing human safety and equity through timely alerts and accessible technology [142], [143], [144], [146], [147].

Historically, the main objective of EWSs has been to quantify the degree of immediate danger that every human settlement in each region of interest is exposed to (see [155], [156], [157], and [158]). In fact, state-of-the-art EWSs used in operational pipelines are based on the analysis of data and measurements at the regional level, which are then aggregated to avoid issues related to the irregularity of the temporal and spatial scales at which each record can be acquired. In this way, state-of-the-art EWSs are prone to highly sensitive and coarse classifications and predictions, potentially leading to a poor perception of risk and inadequate local management plans. Moreover, the classic structure of operational EWSs does not account for cascading effects affecting local communities at social and economic levels. For example, state-of-the-art EWSs for mass movements do not consider a town that can be cut off from transportation and communication systems because of landslides blocking the roads around it as relevant for early warning [158].

AI4EO can be a game changer in this arena [146], [150]. Recent studies [159] have shown the improvement to state-of-the-art EWSs' capacity provided by analysis of the environmental (hydrological and geological) characteristics of mass movements derived from remote sensing platforms and the connectivity information of formal settlements and road network data in terms of graph structures (see Figure 7). This architecture allows us to study the interaction of the probabilistic mass movement susceptibility (derived from the environmental properties by means of a supervised ensemble graph neural network) on a graph representing the communication network (which could involve roads, electricity grids, and water pipelines) connecting the formal settlements. Investigating this structure by graph-based methods (e.g., spectral clustering) results in the derivation of indices to quantify the probability of each human settlement to be directly or indirectly affected by mass movements (i.e., the probability to be hit by a mass movement event—such as a landslide—or the probability to be isolated as a result of mass movements affecting their surroundings) [159]. This approach has the capacity to pave the way for measures and policies for adaptation and mitigation through a new holistic graphical perspective to assess various large-scale geospatial datasets of risk elements, such as exposure, vulnerability, and hazard. In particular, it has been shown that this strategy can deliver robust and reliable outcomes for hindcasting and forecasting of the impact of mass movements at different scales (local, regional, and national), enabling the exploration of long temporal trends (e.g., more than 68,000 incidents of reported mass movements since

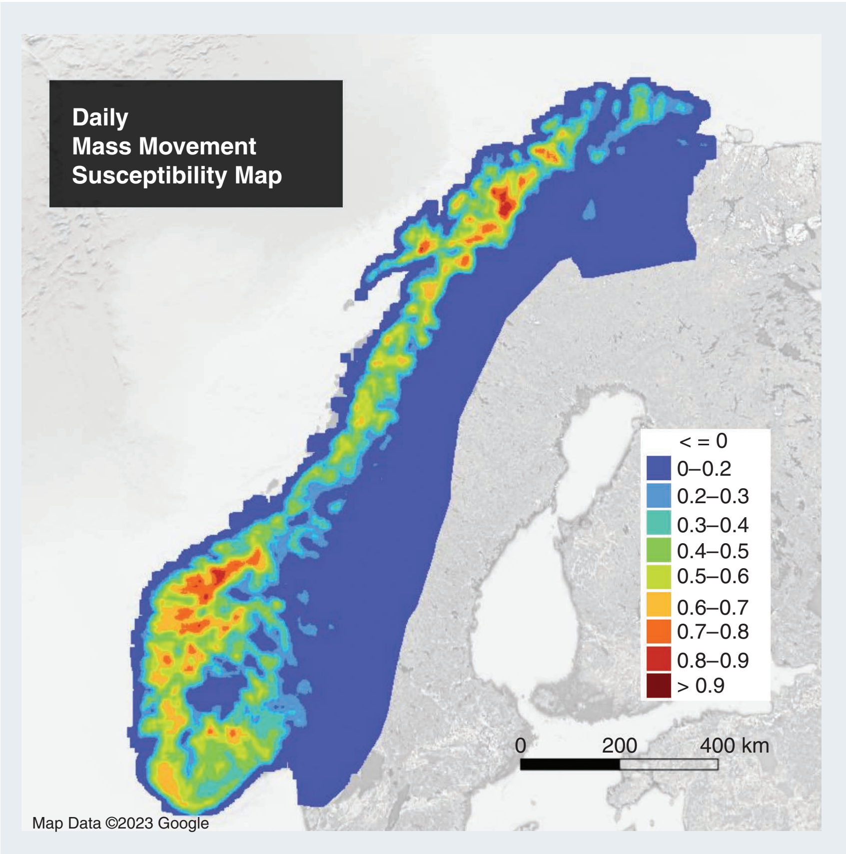

FIGURE 7. Estimation of mass movement susceptibility index across the Norwegian territory by means of the architecture proposed in [159].

1957) and the accurate identification of critical factors for climate change-induced mass movements [159].

CLIMATE TELECONNECTIONS

One of the most important open challenges in environmental monitoring and climate change studies is represented by the analysis of climate teleconnections, that is, the investigation of the interplay between weather phenomena at widely separated locations on Earth. This implies the study of climate patterns that span thousands of kilometers. Examples of teleconnection phenomena include shifts of atmospheric mass and/or pressure back and forth between two distant locations, where changes in atmospheric pressure in one location lead to corresponding changes in pressure in another location far away [160], [161], [162]. Therefore, responsible AI-driven insights are crucial for informing global climate strategies and policies to ensure that all communities benefit equitably.

The water cycle is one of the primary drivers of climate teleconnections worldwide. In this respect, the three world poles (the Arctic, the Antarctic, and high-mountain Asia) play a key role in influencing global weather patterns through their impact on teleconnections [163], [164], [165], [166]. Specifically, the weather and climate of high-mountain Asia and the surrounding region have been demonstrated to be influenced by global modes of variability (MoVs), which are fluctuations of atmospheric and oceanographic variables (e.g., sea surface temperature, rainfall, surface pressure, and wind speed) caused by the continuous exchange of mass, momentum, and energy on all timescales among Earth's atmosphere, oceans, cryosphere, and continental hydrology. For example, the planetary-scale Rossby-wave propagation that causes the intraseasonal variability over central Asia and the northern part of India has been demonstrated to be affected by the Eurasian teleconnection. Also well known are the interplay and impact on high-mountain Asia climate variations of the Indian Ocean Dipole, the El Niño Southern Oscillation, the North Atlantic Oscillation, the central Indian Ocean mode, and the boreal summer intraseasonal oscillation. However, a thorough understanding of the historical behavior of global MoVs and their influence on weather patterns within high-mountain Asia has not yet been derived [160], [161], [163], [164], [165], [166], [167], [168].

The use of advanced technologies for ML, relying on records acquired by EO platforms, plays a key role [160], [162]. In fact, as remote sensors can acquire properties of the previously mentioned environmental variables with precision at a global scale, they enable the investigation of MoVs that drive climate teleconnections worldwide. In particular, when considering high-mountain Asia, the integration of remote sensing data with hydrometeorological climate variables (e.g., the geopotential height at 250 hPa, 2-m air temperature, total precipitation, and fractional snow cover area) leads to a comprehensive exploration of the interactions across several MoVs over high-mountain Asia (e.g., the Eurasian teleconnection, the Indian Ocean Dipole, the North Atlantic Oscillation, and the El Niño Southern Oscillation). This results in the quantification of correlations and cause/effect scenarios across MoVs and climate variables as well as the predictability of weather and climate patterns of the variables of interest, so that proper planning of resources and policies to address and adapt to climate change can be put in place [142], [143], [144], [146], [147], [148], [150].

RESPONSIBLE AI INTEGRATION IN BUSINESS INNOVATION AND SUSTAINABILITY

APPLICATIONS OF RESPONSIBLE AI IN BUSINESS INNOVATION AND SUSTAINABILITY

The integration of AI with low-cost, high-performance computing is fostering business innovation with organizations spanning various sectors and industries, disrupting both commercial and sustainability-focused business models. The integration of responsible AI with satellite data enables businesses to make informed, sustainable decisions that benefit both the environment and their operations. These actions may include monitoring remote infrastructure, assessing climate risk, and dynamically managing supply chains and decision-making processes across various industries. By embracing these advanced technologies, companies can reduce their environmental footprint, enhance resource efficiency, and contribute to the development of more sustainable practices across sectors. Prominent industry players, like Planet [169], Descartes Labs [170], and Maxar Technologies [171], lead this transformation. For example:

- Planet [169] provides daily global satellite sensor imagery to track environmental changes, such as deforestation, carbon emissions, and land degradation. Its platform enables businesses and governments to monitor trends in agriculture, urban development, and environmental shifts worldwide. For example, Planet's collaboration with the World Resources Institute has led to initiatives focused on tracking and reducing global deforestation, helping companies to ensure that their supply chains avoid contributing to environmental harm.
- Descartes Labs [170] aids agricultural and energy sectors by providing insights into soil health, water usage, and crop production. Its AI-driven platform processes satellite imagery to uncover patterns related to crop vitality, water availability, and land management. By leveraging predictive analytics, Descartes Labs empowers farmers to make informed decisions that increase efficiency and reduce the environmental impact of farming.
- Maxar Technologies [171] combines satellite imagery with 3D modeling technology to create highly detailed digital representations of Earth's surface. These advanced tools are applied in urban planning, disaster management, and infrastructure development. Maxar's 3D models help urban planners and construction

companies design energy-efficient and environmentally friendly infrastructure. These solutions promote sustainable city development, enabling businesses to minimize their carbon footprints while supporting cities in building resilient and sustainable urban environments.

CHALLENGES AND OPPORTUNITIES FOR RESPONSIBLE AI ADOPTION

While AI-based solutions offer immense potential, their adoption must be governed by robust safeguards to mitigate risks such as data accessibility inequities, embedded biases, and unintended consequences. For example, farmers may lack direct access to satellite data essential for sustainable agricultural management, while well-funded competitors exploit this information to their advantage. Additionally, AI models without localized agricultural knowledge may lead to overly stringent regulations, disproportionately affecting vulnerable communities. To address these concerns, a focus on governance, open source tools, and equitable standards is essential. This approach ensures fair access to AI benefits, fostering trust and transparency in achieving sustainability goals. Integrating AI into environmental, social, and governance (ESG) frameworks supports risk management and innovation, enabling businesses to reduce carbon footprints, improve biodiversity management, and adhere to ethical practices. This results in reduced carbon footprints, better biodiversity management, and the promotion of ethical practices across industries. ESG reports are generally published yearly by commercial satellite providers, such as Maxar Technologies [171], Planet [169], and Airbus [172], and technology companies, such as Microsoft [173], Google [174], and IBM [175]. Some examples of ESG-oriented use cases of AI4EO include the following:

- *Change monitoring and disaster response*: Satellite data, enhanced by AI, enable businesses and governments to monitor climate change, track deforestation, and assess the aftermath of natural disasters. AI algorithms analyze satellite imagery to detect patterns such as changes in vegetation, carbon emissions, and environmental damage, supporting proactive risk mitigation and sustainable strategy development. Companies can also use these data to report on their sustainability initiatives, ensuring compliance with regulations while reducing environmental impact.
- *Integrating diverse datasets to facilitate comprehensive situation assessment and monitor changes over time*: Microsoft's Planetary Computer [176] combines a multipetabyte catalog of global environmental data that can be queried via application programming interfaces, enabling a scientific environment that allows users to answer global questions about the data and applications that put those answers in the hands of conservation stakeholders.
- *Identifying sources of environmental issues, such as methane emissions from landfills, farms, or pipelines*: Carbon Mapper [177] is a nonprofit organization cofounded by the State of California, NASA's Jet Propulsion Laboratory, Planet, the University of Arizona, Arizona State University, the High Tide Foundation, and RMI that plans to deploy a hyperspectral satellite constellation with the ability to pinpoint, quantify, and track point-source methane and CO_2 emissions.
- *Renewable energy site assessment*: For businesses in the renewable energy sector, AI and satellite data enable more efficient site selection for solar and wind energy projects. By evaluating factors such as sunlight exposure, wind patterns, and topography, AI identifies optimal locations for energy infrastructure, minimizing environmental impact and avoiding damage to sensitive ecosystems. This approach supports the transition to clean energy sources and fosters sustainable development.
- *Sustainable agricultural monitoring*: AI-enhanced satellite data support agricultural businesses by optimizing crop management and promoting precision farming. By analyzing satellite imagery, businesses gain insights into soil health, crop development, and water usage, enabling smarter, more sustainable resource allocation for water, fertilizers, and pesticides.
- *Supply chain optimization and ethical sourcing*: AI-driven satellite data are increasingly used to track supply chain activities and ensure ethical sourcing practices. By examining satellite images, companies can monitor raw material origins and the environmental impact of activities such as mining and forestry. This enables businesses to source materials more sustainably and hold suppliers accountable for their environmental practices.
- *Urban sustainability and smart city development*: Powered by satellite data and AI, businesses are able to design smarter, more sustainable cities. Satellite sensor imagery can be used to monitor urban heat islands, energy consumption, pollution, and sprawl, while AI tools can help design energy-efficient buildings, transportation systems, and waste management strategies. These AI-driven insights help cities optimize energy use, reduce emissions, and enhance climate resilience, contributing to a greener, more livable urban future.
- *Water resource management*: AI and satellite data are essential in optimizing water usage, particularly in industries like agriculture, mining, and manufacturing. By analyzing satellite imagery, businesses can assess water availability and distribution, helping them make better decisions about resource management. This leads to more efficient water use, less waste, and improved preparedness for droughts or floods, ensuring sustainable water management for both businesses and communities.

CONCLUSIONS, REMARKS, AND FUTURE DIRECTIONS

CONCLUSIONS AND REMARKS

This article has addressed the pressing need to navigate the ethical and societal complexities inherent in AI4EO applications. By synthesizing insights from diverse sources, including the academic literature and ethical frameworks, it provides actionable guidance for researchers and practitioners in this rapidly evolving field. The article's significance lies in its comprehensive examination of ethical dimensions, such as privacy, security, fairness, responsibility, and the mitigation of (unfair) biases, shedding light on key considerations that underpin responsible AI practices in EO missions. By fostering collaboration and dialog among stakeholders, the article aims to advance ethical discourse and promote a culture of responsible AI development and deployment. It serves as a foundational framework for navigating ethical dilemmas and guiding the responsible integration of AI technologies into EO missions. Ultimately, the article underscores the transformative potential of AI4EO applications in addressing global challenges while emphasizing the importance of upholding ethical principles to ensure that these technologies serve the collective good and contribute to a more equitable and sustainable world.

Given the current geopolitical state of the world, the role of AI4EO takes on heightened significance in contributing to global stability and peace. Aligning with UN SDG 16, which emphasizes peace, justice, and strong institutions, this work underscores the critical importance of leveraging AI4EO to support these goals and mitigate conflict.

This research article explores the significant effects of AI4EO applications on political, societal, and environmental landscapes. It emphasizes the importance of maintaining high scientific standards and ethical principles to ensure transparency, reproducibility, and trust in AI4EO research. By advocating for open access to data and code, the article facilitates evidence-based decision-making processes in policy making, enabling stakeholders to make well-informed choices. Societally, the article highlights the critical need to address biases and safeguard privacy rights to promote fair outcomes and reduce stigmatization in vulnerable communities. It underscores the importance of responsible AI practices that prioritize fairness and inclusivity, fostering societal trust and cohesion. Environmentally, the article stresses AI's vital role in addressing pressing environmental challenges through sustainable resource management and conservation efforts. By aligning AI4EO workflows with sustainability principles, the article offers a path for leveraging AI technologies to advance the UN SDGs, contributing to a more resilient and sustainable future for humanity and the planet.

FUTURE INVESTIGATIONS

Future investigations, as indicated multiple times throughout this article, should focus on developing strategies that facilitate the open sharing of data for scientific advancements while simultaneously preserving sensitive personal or group information. Ensuring the ethical use and protection of such data is crucial, particularly in applications like mapping deprived urban areas or victim detection, where the responsible use of AI can have significant societal impacts. These applications depend critically on the availability of large benchmark datasets, which are currently lacking. To promote responsible AI, it is essential to implement robust privacy-preserving techniques and data anonymization methods that protect individuals' identities while maintaining the utility of the data for research purposes. This includes using techniques like differential privacy, federated learning, and secure multiparty computation to safeguard sensitive information. Additionally, it is vital to establish clear guidelines and ethical standards for data sharing and usage. These guidelines should ensure that data are used in a manner that respects the rights and dignity of individuals and communities. Collaboration with stakeholders, including ethicists, legal experts, and representatives from affected communities, is necessary to create frameworks that balance the need for scientific progress with the imperative to protect sensitive information. Moreover, transparency in data collection, processing, and sharing practices is essential. Researchers and organizations must be open about the sources of their data, the methods used to process it, and the steps taken to protect privacy. This transparency builds trust and ensures that AI advancements are grounded in ethical practices. By prioritizing the responsible handling of sensitive data, the research community can develop AI-based methods that contribute to societal well-being while respecting individual and group privacy.

LLMs AND NLP HAVE TRANSFORMATIVE POTENTIAL IN ADVANCING RESPONSIBLE AI PRINCIPLES, LIKE DATA PRIVACY, INTEGRITY, AND SECURITY, WITHIN GEOSPATIAL APPLICATIONS.

LLMs and natural language processing (NLP) have transformative potential in advancing responsible AI principles, like data privacy, integrity, and security, within geospatial applications. LLMs can process vast numbers of unstructured data, such as reports, articles, and satellite image annotations, enhancing GeoAI's ability to interpret semantic information about places and events. This leads to more accurate geospatial analysis while maintaining data integrity. Moreover, when designed with privacy considerations, LLMs can identify and handle sensitive geospatial information responsibly, such as by detecting and redacting private location-based data to comply with privacy regulations like the General Data Protection Regulation GDPR. In terms of transparency, LLMs and NLP can make GeoAI models more interpretable by providing human-readable or machine-readable explanations for predictions, where understanding model outputs is critical [178]. Privacy-preserving techniques, such as differential privacy and federated learning [179], can protect sensitive geospatial data. Bias audits and fairness evaluations are essential to prevent unequal representation, particularly in applications like urban planning or disaster relief, where biases could lead to unequal resource allocation. Lastly, adhering to ethical standards and legal regulations ensures that these models operate within legal boundaries, respecting user rights. LLMs and NLP also facilitate interdisciplinary collaboration by translating complex GeoAI outputs into clear summaries, making geospatial insights accessible across fields and supporting responsible data sharing and effective collaboration across environmental, economic, and social domains [140]. We acknowledge that best practices for ensuring the responsible use of these language-driven GeoAI models should be, but are currently not, clearly defined. Therefore, we emphasize the urgent need for further research to establish guiding principles for the responsible use of new multimodal approaches.

In the early 2020s, following decades of model-centric approaches and the deep learning revolution, there was a notable shift from model-centric to data-centric learning. The concept of data-centric AI began gaining traction, emphasizing the quality of data over the complexity of models. Concurrently, a strong societal push emerged with the advent of responsible AI recommendations and legislation, as outlined in this article. This necessitates research to develop methods that can capture biases in compliance with such legislation while maintaining scientific robustness and addressing the unique characteristics of EO data. Responsible AI must consider the entire ML pipeline and involve multiple stakeholders. This includes identifying sensitive attributes and integrating specific domain knowledge on spatial autocorrelation into bias mitigation techniques. Currently, there is a balanced focus on both data quality and model innovation (the era of data–model-centric approaches). Techniques such as foundation models, synthetic data generation, uncertainty modeling and explainability, and robust data validation are employed to enhance data-centric approaches, emphasizing both model and data equally. Ethical considerations, including fairness, transparency, and bias mitigation, are also integral to data-centric AI practices.

ACKNOWLEDGMENT

BPH was supported by grants from the Ministry of Education, Singapore, under its Academic Research Fund Tier3 MOE-T32022-0006. This is Earth Observatory of Singapore (EOS) contribution 639. Work of Weikang Yu was supported by grants from the 3D-ABC project, hosted by the Helmholtz Foundation Model Initiative. The corresponding author is Weikang Yu.

AUTHOR INFORMATION

Pedram Ghamisi (p.ghamisi@gmail.com) is with Helmholtz-Zentrum Dresden-Rossendorf, 09599 Freiberg, Germany, and also with the Lancaster Environment Centre, Lancaster University, LA1 4YR Lancaster, U.K. He is a Senior Member of IEEE.

Weikang Yu (w.yu@tum.de) is with the Chair of Data Science in Earth Observation, Technical University of Munich, Germany, and also with the Helmholtz-Zentrum Dresden-Rossendorf, 09599 Freiberg, Germany. He is a Student Member of IEEE.

Andrea Marinoni (andrea.marinoni@uit.no; am2920@cam.ac.uk) is with UiT - The Arctic University of Norway, 9019 Tromsø, Norway, and the University of Cambridge, CB3 0FD Cambridge, U.K. He is a Senior Member of IEEE.

Caroline M. Gevaert (c.m.gevaert@utwente.nl) is with ITC Faculty Geo-Information Science and Earth Observation, University of Twente, 7522 NH Enschede, The Netherlands.

Claudio Persello (c.persello@utwente.nl) is with ITC Faculty Geo-Information Science and Earth Observation, University of Twente, 7522 NH Enschede, The Netherlands. He is a Senior Member of IEEE.

Sivasakthy Selvakumaran (ss683@cam.ac.uk) is with the University of Cambridge, CB3 0FA Cambridge, U.K.

Manuela Girotto (mgirotto@berkeley.edu) is with the University of California, Berkeley, CA 94720 USA.

Benjamin P. Horton (bphorton@ntu.edu.sg) is with Earth Observatory of Singapore, Nanyang Technological University, Singapore 639798.

Philippe Rufin (philippe.rufin@hu-berlin.de) is with UCLouvain, 1348 Louvain-la-Neuve, Belgium and the Humboldt University of Berlin, 10117 Berlin, Germany.

Patrick Hostert (patrick.hostert@geo.hu-berlin.de) is with the Humboldt University of Berlin, 10117 Berlin, Germany.

Fabio Pacifici is with Maxar Technologies, Westminster, CO 80234 USA, and is the IEEE Geoscience and Remote Sensing Society vice president of technical activities. He is a Senior Member of IEEE.

Peter M. Atkinson (pma@lancaster.ac.uk) is with Lancaster University, LA1 4YR Lancaster, U.K and also with the University of Southampton, SO53 1HB, U.K., and Tongji University, 200092, China.

REFERENCES

[1] P. Ghamisi et al., "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 6, no. 3, pp. 10–43, Sep. 2018, doi: 10.1109/MGRS.2018.2854840.

[2] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.

[3] A. Newell and H. Simon, "The logic theory machine–A complex information processing system," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 61–79, Sep. 1956, doi: 10.1109/TIT.1956.1056797.

[4] D. A. Waterman, *A Guide to Expert Systems*. Reading, MA, USA: Addison-Wesley, 1985.

[5] F. Lehmann, "Semantic networks," *Comput. Math. Appl.*, vol. 23, nos. 2–5, pp. 1–50, Jan./Mar. 1992, doi: 10.1016/0898-1221(92)90135-5.

[6] G. Mountrakis, J. Im, and C. Ogole, “Support vector machines in remote sensing: A review,” *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, May 2011, doi: 10.1016/j.isprsjprs.2010.11.001.

[7] L. Deng, “The MNIST database of handwritten digit images for machine learning research [Best of the Web],” *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012, doi: 10.1109/MSP.2012.2211477.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Piscataway, NJ, USA: IEEE Press, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[9] T.-Y. Lin et al., “Microsoft COCO: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[10] A. Radford, “Improving language understanding by generative pre-training,” *OpenAI Blog*, Jun. 11, 2018. [Online]. Available at: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

[11] A. Radford et al., “Learning transferable visual models from natural language supervision,” in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2021, pp. 8748–8763.

[12] A. Radford et al., “Language models are unsupervised multitask learners,” *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.

[13] C. Schuhmann et al., “Laion-400m: Open dataset of clip-filtered 400 million image-text pairs,” 2021, *arXiv:2111.02114*.

[14] Y. Zhu, O. Groth, M. Bernstein, and L. Fei-Fei, “Visual7W: Grounded question answering in images,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4995–5004, doi: 10.1109/CVPR.2016.540.

[15] K. Marino, M. Rastegari, A. Farhadi, and R. Mottaghi, “OK-VQA: A visual question answering benchmark requiring external knowledge,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3195–3204, doi: 10.1109/CVPR.2019.00331.

[16] J. Buolamwini and T. Gebru, “Gender shades: Intersectional accuracy disparities in commercial gender classification,” in *Proc. Conf. Fairness, Accountability Transparency*, PMLR, 2018, pp. 77–91.

[17] K. Shahriari and M. Shahriari, “IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems,” in *Proc. IEEE Canada Int. Humanitarian Technol. Conf. (IHTC)*, Piscataway, NJ, USA: IEEE Press, 2017, pp. 197–201, doi: 10.1109/IHTC.2017.8058187.

[18] D. Gotterbarn et al., “ACM code of ethics and professional conduct,” 2018. Accessed: Jan. 24, 2025. [Online]. Available at: https://www.acm.org/code-of-ethics

[19] AI HLEG, “Ethics guidelines for trustworthy AI,” B-1049 Brussels, vol. 6, 2019. [Online]. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[20] L. Galindo, K. Perset, and F. Sheeka, “An overview of national ai strategies and policies,” 2021. [Online]. Available at: https://www.oecd.org/en/publications/an-overview-of-national-ai-strategies-and-policies_c05140d9-en.html

[21] G. Mai et al., “On the opportunities and challenges of foundation models for geospatial artificial intelligence,” 2023, *arXiv:2304.06798*.

[22] D. Hong et al., “SpectralGPT: Spectral foundation model,” 2023, *arXiv:2311.07113*.

[23] V. Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, vol. 2156. Cham, Switzerland: Springer-Verlag, 2019.

[24] P. Ghamisi et al., “The potential of machine learning for a more responsible sourcing of critical raw materials,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8971–8988, 2021, doi: 10.1109/JSTARS.2021.3108049.

[25] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, “Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI,” Berkman Klein Center Research Publication, no. 2020-1, 2020. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3518482

[26] B. Friedman and H. Nissenbaum, “Bias in computer systems,” *ACM Trans. Inf. Syst.*, vol. 14, no. 3, pp. 330–347, 1996, doi: 10.1145/230538.230561.

[27] “Recommendation on the ethics of artificial intelligence,” UNESCO, Paris, France, 2021. [Online]. Available: https://www.naavi.org/uploads_wp/2023/Recommendation%20on%20the%20Ethics%20of%20Artificial%20Intelligence%20-%20UNESCO%20Digital%20Library.pdf

[28] “Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts,” Apr. 21, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206

[29] H. Suresh and J. Guttag, “A framework for understanding sources of harm throughout the machine learning life cycle,” in *Proc. Equity Access Algorithms, Mechanisms, Optim.*, 2021, pp. 1–9, doi: 10.1145/3465416.3483305.

[30] T. Hellström, V. Dignum, and S. Bensch, “Bias in machine learning–what is it good for?” 2020, *arXiv:2004.00686*.

[31] S. A. Friedler, C. Scheidegger, S. Venkatasubramanian, S. Choudhary, E. P. Hamilton, and D. Roth, “A comparative study of fairness-enhancing interventions in machine learning,” in *Proc. Conf. Fairness, Accountability, Transparency*, 2019, pp. 329–338, doi: 10.1145/3287560.3287589.

[32] A. Lavell et al., “Climate change: New dimensions in disaster risk, exposure, vulnerability, and resilience,” in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, C. B. Field et al., Eds., Cambridge, U.K.: Cambridge Univ. Press, 2012, pp. 25–64.

[33] P. Lucci, T. Bhatkal, and A. Khan, “Are we underestimating urban poverty?” *World Develop.*, vol. 103, pp. 297–310, Mar. 2018, doi: 10.1016/j.worlddev.2017.10.022.

[34] A. Castelnovo, R. Crupi, G. Greco, D. Regoli, I. G. Penco, and A. C. Cosentini, “A clarification of the nuances in the fairness metrics landscape,” *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 4209, doi: 10.1038/s41598-022-07939-1.

[35] C. M. Gevaert, T. Buunk, and M. J. Van den Homberg, “Auditing geospatial datasets for biases: Using global building datasets for disaster risk management,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 12,579–12,590, 2024, doi: 10.1109/JSTARS.2024.3422503.

[36] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, MA, USA: MIT Press, 2023.

[37] R. Roscher et al., "Data-centric machine learning for geospatial remote sensing data," 2023, *arXiv:2312.05327*.

[38] M. Kuffer, K. Pfeffer, and R. Sliuzas, "Slums from space—15 years of slum mapping using remote sensing," *Remote Sens.*, vol. 8, no. 6, 2016, Art. no. 455, doi: 10.3390/rs8060455.

[39] R. Mahabir, A. Croitoru, A. T. Crooks, P. Agouris, and A. Stefanidis, "A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities," *Urban Sci.*, vol. 2, no. 1, 2018, Art. no. 8, doi: 10.3390/urbansci2010008.

[40] F. Schlosser, V. Sekara, D. Brockmann, and M. Garcia-Herranz, "Biases in human mobility data impact epidemic modeling," 2021, *arXiv:2112.12521*.

[41] N. Karasiak, J.-F. Dejoux, C. Monteil, and D. Sheeren, "Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing," *Mach. Learn.*, vol. 111, no. 7, pp. 2715–2740, 2022, doi: 10.1007/s10994-021-05972-1.

[42] Y. Park et al., "Spatial autocorrelation may bias the risk estimation: An application of eigenvector spatial filtering on the risk of air pollutant on asthma," *Sci. Total Environ.*, vol. 843, Oct. 2022, Art. no. 157053, doi: 10.1016/j.scitotenv.2022.157053.

[43] S. F. E. A. Shaikh et al., "Accounting for spatial autocorrelation is needed to avoid misidentifying trade-offs and bundles among ecosystem services," *Ecol. Indic.*, vol. 129, Oct. 2021, Art. no. 107992, doi: 10.1016/j.ecolind.2021.107992.

[44] W. Yu, Y. Xu, and P. Ghamisi, "Universal adversarial defense in remote sensing based on pre-trained denoising diffusion models," 2023, *arXiv:2307.16865*.

[45] Y. Xu, B. Du, and L. Zhang, "Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: Attacks and defenses," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1604–1617, Feb. 2021, doi: 10.1109/TGRS.2020.2999962.

[46] L. Chen, Z. Xu, Q. Li, J. Peng, S. Wang, and H. Li, "An empirical study of adversarial examples on remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7419–7433, Sep. 2021, doi: 10.1109/TGRS.2021.3051641.

[47] Y. Xu and P. Ghamisi, "Universal adversarial examples in remote sensing: Methodology and benchmark," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, doi: 10.1109/TGRS.2022.3156392.

[48] H. Wu, S. Yunas, S. Rowlands, W. Ruan, and J. Wahlström, "Adversarial driving: Attacking end-to-end autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Piscataway, NJ, USA: IEEE Press, 2023, pp. 1–7, doi: 10.1109/IV55152.2023.10186386.

[49] W. Czaja, N. Fendley, M. Pekala, C. Ratto, and I.-J. Wang, "Adversarial examples in remote sensing," in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2018, pp. 408–411, doi: 10.1145/3274895.3274904.

[50] Y. Zhang et al., "Adversarial patch attack on multi-scale object detection for UAV remote sensing images," *Remote Sens.*, vol. 14, no. 21, 2022, Art. no. 5298, doi: 10.3390/rs14215298.

[51] G. M. Foody and P. M. Atkinson, "Current status of uncertainty issues in remote sensing and GIS," *Uncertainty in Remote Sensing and GIS*, pp. 287–302, 2002.

[52] J. L. Dungan, "Toward a comprehensive view of uncertainty in remote sensing analysis," in *Uncertainty in Remote Sensing and GIS*, G. M. Foody and P. M. Atkinson, Eds., New York, NY, USA: Wiley, 2002, pp. 25–35.

[53] H. J. Persson and G. Ståhl, "Characterizing uncertainty in forest remote sensing studies," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 505, doi: 10.3390/rs12030505.

[54] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 4, no. 2, pp. 41–57, Jun. 2016, doi: 10.1109/MGRS.2016.2548504.

[55] D. Tuia, C. Persello, and L. Bruzzone, "Recent advances in domain adaptation for the classification of remote sensing data," 2021, *arXiv:2104.07778*.

[56] X. Zhang, W. Yu, M.-O. Pun, and W. Shi, "Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 197, pp. 1–17, Mar. 2023, doi: 10.1016/j.isprsjprs.2023.01.018.

[57] M. Xu, M. Wu, K. Chen, C. Zhang, and J. Guo, "The eyes of the gods: A survey of unsupervised domain adaptation methods based on remote sensing data," *Remote Sens.*, vol. 14, no. 17, 2022, Art. no. 4380, doi: 10.3390/rs14174380.

[58] M. Rußwurm, S. Wang, M. Korner, and D. Lobell, "Meta-learning for few-shot land cover classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 200–201, doi: 10.1109/CVPRW50498.2020.00108.

[59] Y. Li, Z. Shao, X. Huang, B. Cai, and S. Peng, "Meta-FSEO: A meta-learning fast adaptation with self-supervised embedding optimization for few-shot remote sensing scene classification," *Remote Sens.*, vol. 13, no. 14, 2021, Art. no. 2776, doi: 10.3390/rs13142776.

[60] O. Tasar, Y. Tarabalka, and P. Alliez, "Incremental learning for semantic segmentation of large-scale remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3524–3537, Sep. 2019, doi: 10.1109/JSTARS.2019.2925416.

[61] W. Liu, X. Nie, B. Zhang, and X. Sun, "Incremental learning with open-set recognition for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, doi: 10.1109/TGRS.2022.3173995.

[62] N. Lin, Y. Chen, H. Liu, and H. Liu, "A comparative study of machine learning models with hyperparameter optimization algorithm for mapping mineral prospectivity," *Minerals*, vol. 11, no. 2, 2021, Art. no. 159, doi: 10.3390/min11020159.

[63] A. Doicu, A. Doicu, D. S. Efremenko, D. Loyola, and T. Trautmann, "An overview of neural network methods for predicting uncertainty in atmospheric remote sensing," *Remote Sens.*, vol. 13, no. 24, 2021, Art. no. 5061, doi: 10.3390/rs13245061.

[64] R. Black et al., “Model risk: Illuminating the black box,” *Brit. Actuarial J.*, vol. 23, 2018, Art. no. e2, doi: 10.1017/S1357321717000150.

[65] W. Brendel, J. Rauber, and M. Bethge, “Decision-based adversarial attacks: Reliable attacks against black-box machine learning models,” 2017, *arXiv:1712.04248*.

[66] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” 2014, *arXiv:1412.6572*.

[67] C. Sitawarin, Z. J. Golan-Strieb, and D. Wagner, “Demystifying the adversarial robustness of random transformation defenses,” in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2022, pp. 20,232–20,252.

[68] Y. Ma, Z. Zhang, Y. Kang, and M. Özdoğan, “Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach,” *Remote Sens. Environ.*, vol. 259, Jun. 2021, Art. no. 112408, doi: 10.1016/j.rse.2021.112408.

[69] G. J. De Lannoy, R. H. Reichle, and J. A. Vrugt, “Uncertainty quantification of GEOS-5 l-band radiative transfer model parameters using Bayesian inference and SMOS observations,” *Remote Sens. Environ.*, vol. 148, pp. 146–157, May 2014, doi: 10.1016/j.rse.2014.03.030.

[70] E. Hasanpour Zaryabi, L. Moradi, B. Kalantar, N. Ueda, and A. A. Halin, “Unboxing the black box of attention mechanisms in remote sensing big data using XAI,” *Remote Sens.*, vol. 14, no. 24, 2022, Art. no. 6254, doi: 10.3390/rs14246254.

[71] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you?” Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144.

[72] A. K. Gizzini, M. Shukor, and A. J. Ghandour, “Extending cam-based XAI methods for remote sensing imagery segmentation,” 2023, *arXiv:2310.01837*.

[73] A. Pal and R. Vidal, “A game theoretic analysis of additive adversarial attacks and defenses,” in *Proc. 34th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2020, pp. 1345–1355.

[74] R. Harris and I. Baumann, “Satellite earth observation and national data regulation,” *Space Policy*, vol. 56, May 2021, Art. no. 101422, doi: 10.1016/j.spacepol.2021.101422.

[75] R. L. Finn, D. Wright, and M. Friedewald, “Seven types of privacy,” in *European Data Protection: Coming of Age*, S. Gutwirth, R. Leenes, P. de Hert, and Y. Poullet, Eds., Dordrecht, The Netherlands: Springer-Verlag, 2013, pp. 3–32.

[76] C. M. Gevaert, R. Sliuzas, C. Persello, and G. Vosselman, “Evaluating the societal impact of using drones to support urban upgrading projects,” *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 3, 2018, Art. no. 91, doi: 10.3390/ijgi7030091.

[77] N. Zhang, F. Nex, G. Vosselman, and N. Kerle, “Training a disaster victim detection network for UAV search and rescue using harmonious composite images,” *Remote Sens.*, vol. 14, no. 13, 2022, Art. no. 2977, doi: 10.3390/rs14132977.

[78] M. Kuffer et al., “The scope of earth-observation to improve the consistency of the SDG slum indicator,” *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 11, 2018, Art. no. 428, doi: 10.3390/ijgi7110428.

[79] M. Kuffer et al., “Spatial information gaps on deprived urban areas (slums) in low-and-middle-income-countries: A user-centered approach,” *Urban Sci.*, vol. 5, no. 4, 2021, Art. no. 72, doi: 10.3390/urbansci5040072.

[80] C. Persello and A. Stein, “Deep fully convolutional networks for the detection of informal settlements in VHR images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2325–2329, Dec. 2017, doi: 10.1109/LGRS.2017.2763738.

[81] M. M. Hassan, I. Hasan, J. Southworth, and T. Loboda, “Mapping fire-impacted refugee camps using the integration of field data and remote sensing approaches,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 115, Dec. 2022, Art. no. 103120, doi: 10.1016/j.jag.2022.103120.

[82] “Kenya slum demolished to make way for road,” BBC News, Jul. 24, 2018. [Online]. Available: https://www.bbc.com/news/world-africa-44931808

[83] G. Chi, H. Fang, S. Chatterjee, and J. E. Blumenstock, “Microestimates of wealth for all low-and middle-income countries,” *Proc. Natl. Acad. Sci.*, vol. 119, no. 3, 2022, Art. no. e2113658119, doi: 10.1073/pnas.2113658119.

[84] B. Hanson et al., “Garbage in, garbage out: Mitigating risks and maximizing benefits of AI in research,” *Nature*, vol. 623, no. 7985, pp. 28–31, 2023, doi: 10.1038/d41586-023-03316-8.

[85] E. Rolf, K. Klemmer, C. Robinson, and H. Kerner, “Mission critical—Satellite data is a distinct modality in machine learning,”2024, *arXiv:2402.01444*.

[86] K. Wagstaff, “Machine learning that matters,” 2012, *arXiv:1206.4656*.

[87] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” *Nature Mach. Intell.*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.

[88] M. Kochupillai, M. Kahl, M. Schmitt, H. Taubenböck, and X. Zhu, “Earth Observation and Artificial Intelligence: Understanding emerging ethical issues and opportunities,” *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 10, no. 4, pp. 90–124, Dec. 2022, doi: 10.1109/MGRS.2022.3208357.

[89] P. Kedron and A. E. Frazier, “How to improve the reproducibility, replicability, and extensibility of remote sensing research,” *Remote Sens.*, vol. 14, no. 21, Oct. 2022, Art. no. 5471, doi: 10.3390/rs14215471.

[90] M. D. Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci. Data*, vol. 3, no. 1, Mar. 2016, Art. no. 160018, doi: 10.1038/sdata.2016.18.

[91] D. B. Richardson, M.-P. Kwan, G. Alter, and J. E. McKendry, “Replication of scientific research: Addressing geoprivacy, confidentiality, and data sharing challenges in geospatial research,” *Ann. GIS*, vol. 21, no. 2, pp. 101–110, Apr. 2015, doi: 10.1080/19475683.2015.1027792.

[92] D. Kinkade and A. Shepherd, “Geoscience data publication: Practices and perspectives on enabling the FAIR guiding principles,” *Geosci. Data J.*, vol. 9, no. 1, pp. 177–186, Jun. 2022, doi: 10.1002/gdj3.120.

[93] M. Kochupillai and H. Taubenböck, “Conducting ethically mindful earth observation research: The case of slum mapping,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE Press, 2023, pp. 1937–1940, doi: 10.1109/IGARSS52108.2023.10281725.

[94] I. Oluoch, “Crossing boundaries: The ethics of AI and geographic information technologies,” *ISPRS Int. J. Geo-Inf.*, vol. 13, no. 3, pp. 87, Mar. 2024, doi: 10.3390/ijgi13030087.

[95] B. Tang, Y. Liu, and D. S. Matteson, “Predicting poverty with vegetation index,” *Appl. Econ. Perspectives Policy*, vol. 44, no. 2, pp. 930–945, 2022, doi: 10.1002/aepp.13221.

[96] A. Branthomme et al., “How much do large-scale and small-scale farming contribute to global deforestation? Results from a remote sensing pilot approach,” FAO, Rome, Italy, Tech. Rep., 2023. [Online]. Available at: https://openknowledge.fao.org/server/api/core/bitstreams/af7565b6-5584-4385-af1e-10a62d1ebe2b/content#:~:text=Deforestation%20due%20to%20cropland%20expansion,percent%20to%20large%2D%20scale%20farming

[97] T. Jayne, A. Wineman, J. Chamberlin, M. Muyanga, and F. K. Yeboah, “Changing farm size distributions and agricultural transformation in sub-Saharan Africa,” *Annu. Rev. Resour. Econ.*, vol. 14, no. 1, pp. 109–130, 2022, doi: 10.1146/annurev-resource-111220-025657.

[98] A. Wineman, T. Jayne, and C. Stevens, “The relationship between medium-scale farms and deforestation in sub-Saharan Africa,” USAID, Washington, DC, USA, 2022. [Online]. Available at: https://www.land-links.org/wp-content/uploads/2022/01/INRM-Activity-7.3-Medium-scale-farms-and-deforestation-FINAL-2021-11-3508.pdf

[99] R. N. Masolele et al., “Mapping the diversity of land uses following deforestation across Africa,” *Sci. Rep.*, vol. 14, no. 1, 2024, Art. no. 1681, doi: 10.1038/s41598-024-52138-9.

[100] C. Chiarella, P. Meyfroidt, D. Abeygunawardane, and P. Conforti, “Balancing the trade-offs between land productivity, labor productivity and labor intensity,” *Ambio*, vol. 52, no. 10, pp. 1618–1634, Oct. 2023, doi: 10.1007/s13280-023-01887-4.

[101] P. Meyfroidt, “Trade-offs between environment and livelihoods: Bridging the global land use and food security discussions,” *Global Food Secur.*, vol. 16, pp. 9–16, Mar. 2018, doi: 10.1016/j.gfs.2017.08.001.

[102] M. G. Tulbure, P. Hostert, T. Kuemmerle, and M. Broich, “Regional matters: On the usefulness of regional land-cover datasets in times of global change,” *Remote Sens. Ecol. Conservation*, vol. 8, no. 3, pp. 272–283, Jun. 2022, doi: 10.1002/rse2.248.

[103] D. Haelewaters, T. A. Hofmann, and A. L. Romero-Olivares, “Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South,” *PLOS Comput. Biol.*, vol. 17, no. 8, 2021, Art. no. e1009277, doi: 10.1371/journal.pcbi.1009277.

[104] R. Schaeffer, B. Miranda, and S. Koyejo, “Are emergent abilities of large language models a mirage?” 2023, *arXiv:2304.15004*.

[105] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, “Self-supervised learning in remote sensing: A review,” 2022, *arXiv:2206.13188*.

[106] J. Jakubik et al., “Foundation models for generalist geospatial artificial intelligence,” 2023, *arXiv:2310.18660*.

[107] R. Manvi, S. Khanna, M. Burke, D. Lobell, and S. Ermon, “Large language models are geographically biased,” 2024, *arXiv:2402.02680*.

[108] R. Bommasani et al., “On the opportunities and risks of foundation models,” 2021, *arXiv:2108.07258*.

[109] G. Doxani et al., “Atmospheric correction inter-comparison eXercise, ACIX-II land: An assessment of atmospheric correction processors for Landsat 8 and Sentinel-2 over land,” *Remote Sens. Environ.*, vol. 285, Feb. 2023, Art. no. 113412, doi: 10.1016/j.rse.2022.113412.

[110] S. Skakun et al., “Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2,” *Remote Sens. Environ.*, vol. 274, Jun. 2022, Art. no. 112990, doi: 10.1016/j.rse.2022.112990.

[111] P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, “Good practices for estimating area and assessing accuracy of land change,” *Remote Sens. Environ.*, vol. 148, pp. 42–57, May 2014, doi: 10.1016/j.rse.2014.02.015.

[112] A. R. Ives, L. Zhu, F. Wang, J. Zhu, C. J. Morrow, and V. C. Radeloff, “Statistical inference for trends in spatiotemporal data,” *Remote Sens. Environ.*, vol. 266, Dec. 2021, Art. no. 112678, doi: 10.1016/j.rse.2021.112678.

[113] D. Frantz et al., “Understanding the robustness of spectral-temporal metrics across the global Landsat archive from 1984 to 2019 – A quantitative evaluation,” *Remote Sens. Environ.*, vol. 298, Dec. 2023, Art. no. 113823, doi: 10.1016/j.rse.2023.113823.

[114] Y. Xu, T. Bai, W. Yu, S. Chang, P. M. Atkinson, and P. Ghamisi, “AI security for geoscience and remote sensing: Challenges and future trends,” *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 11, no. 2, pp. 60–85, Jun. 2023, doi: 10.1109/MGRS.2023.3272825.

[115] T. Brandsen and M. Honingh, “Definitions of co-production and co-creation,” in *Co-production and co-creation*, T. Brandsen, T. Steen, and B. Verschuere, Eds., Milton Park, U.K. Routledge, 2018, pp. 9–17.

[116] P. H. Verburg et al., “Land system science and sustainable development of the earth system: A global land project perspective,” *Anthropocene*, vol. 12, pp. 29–41, Dec. 2015, doi: 10.1016/j.ancene.2015.09.004.

[117] S. Fritz et al., “Geo-Wiki: An online platform for improving global land cover,” *Environ. Modell. Software*, vol. 31, pp. 110–123, May 2012, doi: 10.1016/j.envsoft.2011.11.015.

[118] D. Schepaschenko et al., “Development of a global hybrid forest mask through the synergy of remote sensing, crowdsourcing and FAO statistics,” *Remote Sens. Environ.*, vol. 162, pp. 208–220, Jun. 2015, doi: 10.1016/j.rse.2015.02.011.

[119] C. M. Gevaert, “Explainable AI for earth observation: A review including societal and regulatory perspectives,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102869, doi: 10.1016/j.jag.2022.102869.

[120] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[121] M. Graziani et al., “A global taxonomy of interpretable AI: Unifying the terminology for the technical and social sciences,” *Artif. Intell. Rev.*, vol. 56, no. 4, pp. 3473–3504, Apr. 2023, doi: 10.1007/s10462-022-10256-8.

[122] C. Nakalembe and H. Kerner, “Considerations for AI-EO for agriculture in Sub-Saharan Africa,” *Environ. Res. Lett.*, vol. 18, no. 4, Mar. 2023, Art. no. 041002, doi: 10.1088/1748-9326/acc476.

[123] B. C. Stahl, D. Schroeder, and R. Rodrigues, *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*. Cham, Switzerland: Springer-Verlag, 2023.

[124] D. Frantz, "FORCE—Landsat + Sentinel-2 analysis ready data and beyond," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1124, doi: 10.3390/rs11091124.

[125] H. J. Miller, "Tobler's first law and spatial analysis," *Ann. Assoc. Amer. Geographers*, vol. 94, no. 2, pp. 284–289, Jun. 2004, doi: 10.1111/j.1467-8306.2004.09402005.x.

[126] D. Frantz, P. Hostert, P. Rufin, S. Ernst, A. Röder, and S. van der Linden, "Revisiting the past: Replicability of a historic long-term vegetation dynamics assessment in the era of big data analytics," *Remote Sens.*, vol. 14, no. 3, pp. 597–27, Jan. 2022, doi: 10.3390/rs14030597.

[127] L. Blickensdörfer, K. Oehmichen, D. Pflugmacher, B. Kleinschmit, and P. Hostert, "National tree species mapping using Sentinel-1/2 time series and German National Forest Inventory data," *Remote Sens. Environ.*, vol. 304, Apr. 2024, Art. no. 114069, doi: 10.1016/j.rse.2024.114069.

[128] F. Lobert et al., "A deep learning approach for deriving winter wheat phenology from optical and SAR time series at field level," *Remote Sens. Environ.*, vol. 298, Dec. 2023, Art. no. 113800, doi: 10.1016/j.rse.2023.113800.

[129] M. Schwieder et al., "Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112795, doi: 10.1016/j.rse.2021.112795.

[130] M. A. Wulder et al., "Virtual constellations for global terrestrial monitoring," *Remote Sens. Environ.*, vol. 170, pp. 62–76, Dec. 2015, doi: 10.1016/j.rse.2015.09.001.

[131] P. Rufin, S. Wang, S. N. Lisboa, J. Hemmerling, M. G. Tulbure, and P. Meyfroidt, "Taking it further: Leveraging pseudo-labels for field delineation across label-scarce smallholder regions," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 34, Nov. 2024, Art. no. 104149, doi: 10.1016/j.jag.2024.104149.

[132] S. Wang, F. Waldner, and D. B. Lobell, "Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision," *Remote Sens.*, vol. 14, no. 22, Jan. 2022, Art. no. 5738, doi: 10.3390/rs14225738.

[133] D. Tuia, R. Roscher, J. D. Wegner, N. Jacobs, X. Zhu, and G. Camps-Valls, "Toward a collective agenda on AI for Earth science data analysis," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 9, no. 2, pp. 88–104, Jun. 2021, doi: 10.1109/MGRS.2020.3043504.

[134] D. Tuia et al., "Artificial intelligence to advance Earth observation: A perspective," 2023, *arXiv:2305.08413*.

[135] "Transforming our world: The 2030 agenda for sustainable development," UN, New York, NY, USA, 2015. [Online]. Available at: https://sdgs.un.org/2030agenda

[136] C. Persello et al., "Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 10, no. 2, pp. 172–200, Jun. 2022, doi: 10.1109/MGRS.2021.3136100.

[137] R. Vinuesa et al., "The role of artificial intelligence in achieving the Sustainable Development Goals," *Nature Commun.*, vol. 11, no. 1, Jan. 2020, Art. no. 233, doi: 10.1038/s41467-019-14108-y.

[138] R. Wilkinson et al., "Environmental impacts of earth observation data in the constellation and cloud computing era," *Sci. Total Environ.*, vol. 909, Jan. 2024, Art. no. 168584, doi: 10.1016/j.scitotenv.2023.168584.

[139] E. Benami et al., "Uniting remote sensing, crop modelling and economics for agricultural risk management," *Nature Rev. Earth Environ.*, vol. 2, no. 2, pp. 140–159, Jan. 2021, doi: 10.1038/s43017-020-00122-y.

[140] Y. Wang, W. Yu, M. Kopp, and P. Ghamisi, "ChangeMinds: Multi-task framework for detecting and describing changes in remote sensing," 2024, *arXiv:2410.10047*.

[141] M. Burke, A. Driscoll, D. B. Lobell, and S. Ermon, "Using satellite imagery to understand and promote sustainable development," *Science*, vol. 371, no. 6535, 2021, Art. no. eabe8628, doi: 10.1126/science.abe8628.

[142] H.-O. Portner et al., "Climate change 2022: Impacts, adaptation and vulnerability - Working group II to the sixth assessment report of the intergovernmental panel on climate change," Cambridge Univ. Press, Cambridge, U.K., 2023. [Online]. Available at: https://www.ipcc.ch/report/sixth-assessment-report-working-group-ii/

[143] V. Masson-Delmotte et al., "Climate change 2021: The physical science basis - Working group I to the sixth assessment report of the intergovernmental panel on climate change," Cambridge Univ. Press, Cambridge, U.K., 2022. [Online]. Available at: https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/

[144] S. Werners et al., "Adaptation pathways: A review of approaches and a learning framework," *Environ. Sci. Policy*, vol. 116, pp. 266–275, Feb. 2021, doi: 10.1016/j.envsci.2020.11.003.

[145] D. Rolnick et al., "Tackling climate change with machine learning," *ACM Comput. Surv.*, vol. 55, no. 2, pp. 1–96, 2022, doi: 10.1145/3485128.

[146] ECMWF, "ECMWF strategy 2021-2030," ECMWF, Reading, U.K., 2021. [Online]. Available: https://www.ecmwf.int/node/19880

[147] Constanze Fetting, "The European green deal," ESDN, 2020. [Online]. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

[148] Ministry of Sustainability and the Environment, "Greengov.sg: Report for financial year 2022," Singapore Green Plan, 2023. [Online]. Available at: https://isomer-user-content.by.gov.sg/23/7a9882e1-94a0-43a8-b4da-7ea2a3c39953/greengovsg-report-fy2022.pdf

[149] 117th Congress of the United States of America, Inflation Reduction Act, Public Law 117-169, vol. 136, 2022, pp. 1818–2090.

[150] P. Griffiths et al., "The ESA green transition information factories? Using earth observation and cloud-based analytics to address the green transition information needs," in *Proc. EGU General Assem.*, Vienna, Austria, 2023. [Online]. Available at: https://meetingorganizer.copernicus.org/EGU23/EGU23-5862.html

[151] M.-H. Lo and J. Famiglietti, "Irrigation in California's central valley strengthens the southwestern U.S. water cycle," *Geophys. Res. Lett.*, vol. 40, no. 2, pp. 301–306, 2013, doi: 10.1002/grl.50108.

[152] L. Pei, N. Moore, S. Zhong, A. D. Kendall, Z. Gao, and D. W. Hyndman, "Effects of irrigation on summer precipitation over the United States," *J. Clim.*, vol. 29, no. 10, pp. 3541–3558, 2016, doi: 10.1175/JCLI-D-15-0337.1.

[153] A. DeAngelis et al., "Evidence of enhanced precipitation due to irrigation over the great plains of the united states," *J. Geophys. Res. Atmos.*, vol. 115, no. D15, 2010, doi: 10.1029/2010JD013892. [Online]. Available at: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2010JD013892

[154] J. Wei, P. A. Dirmeyer, D. Wisser, M. G. Bosilovich, and D. M. Mocko, "Where does the irrigation water go? An estimate of the contribution of irrigation to precipitation using MERRA," *J. Hydrometeorol.*, vol. 14, no. 1, pp. 275–289, 2016, doi: 10.1175/JHM-D-12-079.1.

[155] M. Azarafza, M. Azarafza, H. Akgün, P. M. Atkinson, and R. Derakhshani, "Deep learning-based landslide susceptibility mapping," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 24112, doi: 10.1038/s41598-021-03585-1.

[156] J.-S. L'Heureux, B. K. F. Bache, S. Hov, and S. Lacasse, "Mitigation measures following landslides in sensitive clays, examples from Norway," in *Proc. Geohazards*, vol. VIII, 2022. [Online]. Available at: https://www.researchgate.net/publication/366513460_Mitigation_measures_following_landslides_in_sensitive_clays_-_Examples_from_Norway

[157] C. Jaedicke, K. Lied, and K. Kronholm, "Integrated database for rapid mass movements in Norway," *Nat. Hazards Earth Syst. Sci.*, vol. 9, no. 2, pp. 469–479, 2009, doi: 10.5194/nhess-9-469-2009.

[158] A. Solheim, K. Sverdrup-Thygeson, and B. Kalsnes, "Hazard and risk assessment for early phase road planning in Norway," *Nat. Hazards*, vol. 119, no. 2, pp. 943–963, 2023, doi: 10.1007/s11069-022-05729-6.

[159] J. Dimasaka, S. Selvakumaran, and A. Marinoni, "Near-real-time country-wide estimation of susceptibility and settlement exposure from Norwegian mass movements via inter-graph representation learning," in *Proc. AGU General Assem.*, San Francisco, CA, 2023. [Online]. Available at: https://essopenarchive.org/doi/full/10.22541/essoar.170914538.81935637

[160] E. Massoud, Y.-K. Lim, L. Andrews, and M. Girotto, "Connecting global modes of variability to climate in high mountain Asia," *Atmosphere*, vol. 15, no. 2, 2024, Art. no. 142, doi: 10.3390/atmos15020142.

[161] J. Li, R. Yu, T. Zhou, and B. Wang, "Why is there an early spring cooling shift downstream of the Tibetan Plateau?" *J. Clim.*, vol. 18, no. 22, pp. 4660–4668, 2005, doi: 10.1175/JCLI3568.1.

[162] E. Massoud et al., "Seasonal forecasting skill for the high mountain Asia region in the Goddard Earth observing system," *Earth Syst. Dyn.*, vol. 14, no. 1, pp. 147–171, 2023, doi: 10.5194/esd-14-147-2023.

[163] W. Immerzeel et al., "Importance and vulnerability of the world's water towers," *Nature*, vol. 577, no. 7790, pp. 364–369, 2020, doi: 10.1038/s41586-019-1822-y.

[164] D. Nash, L. M. V. Carvalho, C. Jones, and Q. Ding, "Winter and spring atmospheric rivers in high mountain Asia: Climatology, dynamics, and variability," *Clim. Dyn.*, vol. 58, nos. 9–10, pp. 2309–2331, 2022, doi: 10.1007/s00382-021-06008-z.

[165] Q. Ding and B. Wang, "Intraseasonal teleconnection between the summer Eurasian wave train and the Indian monsoon," *J. Clim.*, vol. 20, no. 15, pp. 3751–3767, 2007, doi: 10.1175/JCLI4221.1.

[166] Y.-K. Lim, "The East Atlantic/West Russia (EA/WR) teleconnection in the North Atlantic: Climate impact and relation to Rossby wave propagation," *Clim. Dyn.*, vol. 44, nos. 11–12, pp. 3211–3222, 2015, doi: 10.1007/s00382-014-2381-4.

[167] M. Stuecker et al., "Revisiting ENSO/Indian ocean dipole phase relationships," *Geophys. Res. Lett.*, vol. 44, no. 5, pp. 2481–2492, 2017, doi: 10.1002/2016GL072308.

[168] Y.-F. Sang, V. Singh, and K. Xu, "Evolution of IOD-ENSO relationship at multiple time scales," *Theor. Appl. Climatol.*, vol. 136, nos. 3–4, pp. 1303–1309, 2019, doi: 10.1007/s00704-018-2557-7.

[169] "Planet labs PBC: ESG report 2023," 2023. Accessed: May 22, 2024. [Online]. Available: https://assets.planet.com/docs/Planet_Labs_PBC_2023_ESG_Report.pdf

[170] "Earth observation geospatial data analytics for insurance." Descartes Labs. Accessed: May 21, 2024. [Online]. Available: https://descarteslabs.com/products/ascend

[171] "Maxar publishes 2022 ESG report," Maxar Technologies, Westminster, CO, USA, 2022. Accessed: May 21, 2024. [Online]. Available: https://www.maxar.com/press-releases/maxar-publishes-2022-esg-report

[172] "Sustainability." Airbus. Accessed: May 21, 2024. [Online]. Available: https://www.airbus.com/en/sustainability

[173] "Ai for good lab." Microsoft. Accessed: May 21, 2024. [Online]. Available: https://www.microsoft.com/en-us/research/group/ai-for-good-research-lab/

[174] "Sustainable innovation and technology." Google Sustainability. Accessed: May 21, 2024. [Online]. Available: https://sustainability.google/

[175] IBM AI Ethics Board, "For the planet and people: IBM'S focus on AI ethics in sustainability," IBM, Apr. 22, 2024. Accessed: May 21, 2024. [Online]. Available: https://www.ibm.com/blog/for-the-planet-and-people-ibms-focus-on-ai-ethics-in-sustainability/

[176] "Microsoft planetary computer." Microsoft. Accessed: May 21, 2024. [Online]. Available: https://planetarycomputer.microsoft.com/

[177] "Carbon mapper launches satellite program to pinpoint methane." Planet. Accessed: May 21, 2024. [Online]. Available: https://www.planet.com/pulse/carbon-mapper-launches-satellite-program-to-pinpoint-methane-and-co2-super-emitters/

[178] A. C. Roman et al., "Open datasheets: Machine-readable documentation for open datasets and responsible AI assessments," 2024, *arXiv:2312.06153*.

[179] B. Buyuktas, G. Sumbul, and B. Demir, "Federated learning across decentralized and unshared archives for remote sensing image classification: A review," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 12, no. 3, pp. 64–80, Sep. 2024, doi: 10.1109/MGRS.2024.3415391.

GRS