News & views

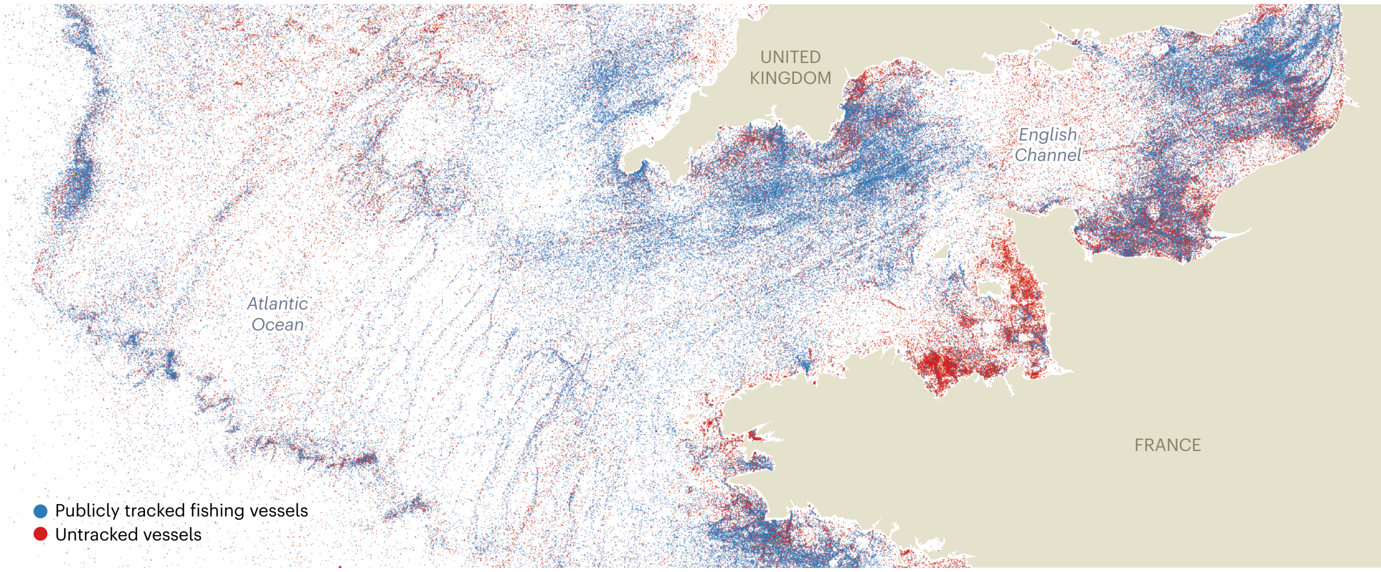

Figure 1 | Industrial fishing around northwest France and the southwest of the United Kingdom. Paolo *et al.*[1] used machine learning to analyse satellite imagery of the oceans taken from 2017 to 2021, and thereby to identify infrastructure associated with industrial activities at sea (such as ships and wind turbines). Each dot represents an individual fishing vessel; blue dots are vessels that were also tracked on public systems, whereas red dots represent previously untracked vessels. (Adapted from Extended Data Fig. 3 of ref. 1.)

Environmental science

Satellites reveal scale of humans' ocean activities

Konstantin Klemmer & Esther Rolf

Machine learning and satellite imagery have been used to map industrial infrastructure at sea — from fishing vessels to wind turbines. The findings provide a more comprehensive picture of maritime activity than ever before. **See p.85**

More than 70% of Earth's surface is covered by oceans, which provide food for billions of people and are used for transporting goods around the planet. Offshore wind, oil and gas infrastructure is also crucial for global energy supplies. Monitoring human activity at sea has been challenging, because of the inaccessibility and large extents of the regions to be monitored. On page 85, Paolo *et al.*[1] show that satellite data can be used to map the human presence on the oceans to an unprecedented extent. The findings reveal that a substantial amount of human activity is not publicly tracked.

Human activities at sea involve the use of stationary structures, such as oil rigs, and movable objects such as fishing vessels. Until now, no comprehensive, global map of these different types of maritime infrastructure had been available. Automated identification systems (AISs) can track maritime traffic using transponder signals emitted from ships, but might not provide a complete picture — either because vessels have poor connectivity to the AIS or because tracking is intentionally disabled.

Paolo *et al.* report an extensive mapping effort that provides a more complete quantification of human activities in the oceans than ever before (see Fig. 1). The authors used deep learning to probe satellite imagery from two sources to detect both stationary and movable objects, mostly in coastal waters, across the global ocean from 2017 to 2021. The authors estimate that 72–76% of all industrial fishing vessels, and 21–30% of transport and energy vessels, are not accounted for by AISs. The findings show that estimates based on previously available, incomplete data massively underestimated the extent of fishing operations in regions such as South and south-east Asia.

The satellite data also allowed changes in activity to be tracked over time. For example, progress in transitioning to the use of clean energy was documented by the appearance of offshore wind turbines. The number of turbines doubled globally during the study period, increasing enormously in some regions (by 900% in China, for example, where about 950 turbines were installed per year, on average). The data also show the effect of the COVID-19 pandemic on global fishing operations: fishing decreased by about 12% at the start of the pandemic, and had still not returned to pre-pandemic levels by the end of 2021. Such data could aid the analysis of the effects of events such as pandemics or natural disasters, and inform maritime policy.

Paolo and colleagues' study highlights the broader implications of automated Earth observation and its increasing capacity to map human activity at a global scale. First, there are direct implications for private- and public-sector decision makers. For example, tracking of clandestine vessels could boost the detection of illegal fishing operations and measurement of the informal economy[2] (economic activities that are neither taxed nor monitored by governments). Indeed, the findings reveal a high concentration of illicit fishing to the west of the Korean Peninsula and on the North African coast. Vessel tracking could also transform environmental conservation efforts by revealing encroachment on protected areas — Paolo *et al.* report that two such areas, the Galapagos Marine Reserve and the Great Barrier Reef Marine Park, were visited by an average of more than 5 and more than 20 fishing vessels per week, respectively.

Second, a lot of global satellite imagery is freely available to the public, and machine-learning tools for processing these data are continuously being developed, often as open-source software (Paolo and colleagues' data set and software are also freely available). Such efforts democratize access to data and tools and allow researchers, analysts and policymakers in low-income countries to leverage tracking technologies at low cost, for example to monitor exclusive economic zones — areas of the ocean for which sovereign states have exclusive exploration and usage rights.

The study also highlights some general limitations of Earth observation by satellites. The spatial resolution of publicly available satellite imagery (such as the Sentinel-1 data used by Paolo *et al.*, which have a ground resolution of about 20 metres per pixel) prevents the detection of objects such as small fishing vessels. Crucially, this leaves many small-scale fishers in coastal communities off the map. And although satellites cover every corner of the planet, some areas are more difficult to map than others — for example, because of persistent cloud coverage.

Emerging data sources and increasingly powerful machine-learning systems will overcome some of these limitations, and thereby enable the quantification of previously unmeasured aspects of human development and change. High-resolution radar instruments can see through clouds, for instance; and advances in 'unsupervised' deep learning will make it possible to gain insight from satellite images with less laborious hand-labelling by humans than is currently needed[3].

The complexity and sheer volume of Earth-observation data remain challenges that are driving research among machine-learning and Earth scientists[4]. But this should not be the only research focus. Earth-observation technologies should also be designed to help capture human activity in a way that is equitable and community-centred. For example, unlike Western large-scale farming operations, many sustenance farmers in low-income countries operate on small plots of land. Earth-observation models can fail to delineate those fields if the imagery available is not of sufficiently high resolution or because models trained on Western data have not been exposed to such small-scale patterns.

Moreover, cross-disciplinary expertise and the involvement of multiple stakeholders are often essential to interpret satellite-based observations in specific regions and put them into context. As an illustration of this, Indigenous peoples often have a fundamentally different, and frequently more complete, understanding of their local ecosystems than can be captured on a satellite image. Failure to incorporate those insights might lead to the development of methods that fail to meet real-world needs, ignore important equity and transparency considerations, and ultimately limit the utility of satellite-based observations[5]. The development of approaches that prioritize computational accessibility — publicly available data, open-source software, and efficient algorithms that don't require high-performance computers — will also be crucial[6].

Paolo and colleagues' study adds to a growing body of work highlighting how deep learning can facilitate ocean monitoring, with applications ranging from the detection of marine debris[7] to tracking algal blooms[8]. The task now is for interdisciplinary collaborations to build on these prototype technologies to set up large-scale observation systems that focus on stakeholder engagement and local community efforts. This will ensure that advances in deep learning for Earth observation achieve their potential to address pressing local and global challenges.

Konstantin Klemmer is at Microsoft Research New England in Cambridge, Massachusetts 02142, USA. **Esther Rolf** is at the Harvard Data Science Initiative and the Center for Research on Computation and Society, Harvard University, Boston, Massachusetts 02134, USA.
e-mails: kklemmer@microsoft.com; erolf@g.harvard.edu

1. Paolo, F. *et al. Nature* **625**, 85–91 (2024).
2. Angrist, N., Goldberg, P. K. & Jolliffe, D. *J. Econ. Persp.* **35**, 215–242 (2021).
3. Wang, Y., Albrecht, C. M., Braham, N. A. A., Mou, L. & Zhou, X. X. *IEEE Geosci. Remote Sensing Mag.* **10**, 213–247 (2022).
4. Reichstein, M. *et al. Nature* **566**, 195–204 (2019).
5. De-Arteaga, M., Herlands, W., Neill, D. B. & Dubrawski, A. *ACM Trans. Mgmt Inf. Syst.* **9**, 9 (2018).
6. Rolf, E. *et al. Nature Commun.* **12**, 4392 (2021).
7. Rußwurm, M., Venkatesa, S. J. & Tuia, D. *iScience* **26**, 108402 (2023).
8. Wang, M. & Hu, C. *Remote Sens. Environ.* **264**, 112631 (2021).

The authors declare no competing interests.

Human genetics

Linking the non-coding genome to human health

Ryan S. Dhindsa & Slavé Petrovski

An expanded version of a human-genome database called gnomAD, containing 76,156 whole-genome sequences, has enabled investigation of how variants in non-protein-coding regions of the genome affect health. **See p.92**

Scientists have long suspected that many disease-causing genetic mutations reside in the 98% of the genome that does not encode proteins, especially in regions that have roles in regulating gene expression. However, it has been challenging to differentiate systematically between harmful and neutral mutations, partly because researchers lack a clear picture of which stretches of the non-coding genome are essential for human health. On page 92, Chen *et al.*[1] address this challenge, introducing a tool that can analyse large collections of human genomes to identify non-coding regions that have the greatest potential to cause disease when mutated.

This work represents the most recent iteration of the Genome Aggregation Database (gnomAD), a publicly available catalogue of human genetic variation. The first version[2], released in 2020, included sequence data from the protein-coding DNA of 125,748 people and the whole genomes of 15,708 people. Since then, the consortium has greatly expanded the database; the resource now includes whole-genome sequences from 76,156 individuals of diverse ancestries, providing a much deeper picture of human genetic variation.

GnomAD has transformed human genetics, especially in terms of diagnosing rare diseases. The genome of any individual differs from those of other people at millions of sites. Most of these genetic variants are clinically insignificant, particularly those that are common in the general population. When clinical