



BIOASSEMBLER

Integrating bio-inspired assembly
into semiconductor manufacturing
technology for biosensors

Forecast on Using Big Data & AI with Multiplexed Silicon Biosensors

(Project N:o 101070589)

Version: V2.0

Status: Final

Date: 29 August 2025

Project Coordinator: VTT Technical
Research Centre of Finland

Project Start Date: 1st September 2022

Project Duration: 42 months

Deliverable N°: D6.4

Dissemination level: Public

Authors: Aino-Lotta Alahäivälä

Contact Person:

Aino-Lotta Alahäivälä, Health Data
Analytics Team, VTT Technical Research
Centre of Finland, aino.alahaivala@vtt.fi



**Funded by
the European Union**

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070589. This document reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein

Revision history

Version	Status	Date	Contributor / Partner	Summary of Changes
1.0	Draft 1	30.06.2025	Aino-Lotta Alahäivälä	Drafting Report
2.0	Finalization of report	29.08.2025	Aino-Lotta Alahäivälä	Review and integration

Contact info

Project Coordinator: Petri Saviranta, VTT Technical Research Centre of Finland,

petri.saviranta@vtt.fi

Work Package 6 (Testing Validation & Data Analysis) Coordinator: Sami Lähteenmäki, VTT Technical Research Centre of Finland, sami.lahteenmaki@vtt.fi

Email: info@bioassembler.eu

List of Contents

Summary	4
1. Introduction	5
2. Literature Survey	7
2.1. Significance, Forecasting & Modelling	7
2.1.1. Advancements in Point-of-Care Diagnostics	7
2.1.2. Contributions to AI & Predictive Modelling	8
2.1.3. Bio-Inspired Manufacturing Innovations.....	9
3.1. Health Data Collection	9
3.2. Health Data Protection & Ethical Considerations	11
3.3. Environmental Measurement Data	14
3.4. Data Integration.....	16
3.5. Big Data Analysis.....	18
3.6. Artificial Intelligence (AI)	21
3.7. Use Cases & Applications.....	24
3.7.1. Disease Outbreak Tracking.....	24
3.7.2. Precision Medicine	25
3.7.3. Environmental Exposure and Public Health	25
3.7.4. Agriculture and Food Quality	26
3.7.5. Data Flow and Oversight: From Personal to Population-Level Insights	26
4. Conclusion	28
Key Recommendations	29
References.....	31

List of figures

Figure 1 – Concept draft.	27
--------------------------------	----

Summary

This document examines the prospective of integrating Big Data and artificial intelligence (AI) with multiplexed silicon biosensors, focusing on their role in health and environment monitoring solutions. Multiplexed silicon biosensors are an emerging class of miniaturized diagnostic devices capable of detecting multiple biological or chemical analytes simultaneously. Coupling these biosensors with Big Data infrastructure and AI analytics promises transformative advances in both healthcare and environmental monitoring. Real-time data streams from wearable and point-of-care biosensors can be aggregated into large databases, allowing AI algorithms to detect subtle patterns and correlations across thousands of sensor readings—yielding observations that human eye might miss. For example, AI-enhanced analysis of a multiplex panel of biomarkers can improve disease diagnosis accuracy compared to single-marker tests, and networks of environmental sensors can feed predictive models to forecast pollution or outbreak trends.

This report surveys the state-of-the-art technologies enabling such integrations, explores current use cases in health and environment, and discusses the technical, regulatory, and ethical challenges that must be addressed to realize these opportunities. Key challenges include ensuring data quality and interoperability, protecting personal data under stringent privacy laws (e.g., General Data Protection Regulation, GDPR), and maintaining public trust through transparency and accountability. The European policy landscape—including the GDPR and forthcoming frameworks like the European Health Data Space and AI Act—provides a blueprint for balancing innovation with ethical safeguards. Recommendations emphasize the need for common data standards, inclusive access to biosensor technology, continuous ethical oversight (such as algorithm audits), user-centric control over personal data assets, and further research into privacy-preserving data analysis techniques. With a human-centric, policy-aware approach, the integration of Big Data and AI with multiplexed silicon biosensors could usher in more proactive health management, rapid disease outbreak detection, and smarter environmental stewardship, all while upholding fundamental rights and public trust.

Keywords: Multiplexed Biosensors, Artificial Intelligence, Big Data Analytics, European Health Data Space, Data Governance and Ethics

1. Introduction

Silicon-based multiplex biosensors are integrated diagnostic devices fabricated using semiconductor microtechnology that can simultaneously detect multiple biological or chemical targets. In essence, a biosensor consists of a selective biorecognition element (such as an enzyme, antibody, or aptamer that binds a specific analyte) coupled to a transducer that converts that biochemical interaction into a measurable signal. Silicon provides an ideal platform for miniaturizing these components: modern silicon photonic and microelectronic biosensors offer high sensitivity and the ability to integrate numerous sensing elements on a single chip for parallel (multiplexed) analyte detection (Jiménez-Rodríguez et al., 2022). Such chips are CMOS-compatible and can be mass-produced at low cost and small size, enabling portable point-of-care (POC) devices that bring sophisticated lab testing directly to patients or field settings (Bouderhem, 2024; Jiménez-Rodríguez et al., 2022). For example, a silicon photonic biosensor array might incorporate many microring resonators on one chip, each functionalized for a different biomarker, allowing one small device to test for a panel of conditions simultaneously. In operation, a multiplexed biosensor could accept a tiny fluid sample (blood, saliva, etc.), pass it across the 1 array of micro-sensors, and within minutes provide readings on several biomarkers at once in a handheld format.

These multiplex biosensors are highly relevant to POC diagnostics because they enable rapid on-site testing at the bedside, clinic, or even at home, instead of sending samples to central labs and waiting days for results. This shift produces rich datasets in real time, which can be immediately used for data analysis and decision-making (Jiménez-Rodríguez et al., 2022). With the integration of wireless electronics, many silicon biosensor chips now include on-board signal processing and communication modules to transmit data to computers or smartphones. Connecting biosensors with digital networks allows automatic data logging and remote monitoring: each biosensor reading becomes a machine-readable data point that can be integrated with electronic health records (EHRs) and aggregated into large databases. This connectivity enables the application of Big Data analytics—detecting patterns across large-scale sensor datasets, correlating sensor data with clinical or environmental outcomes, and applying AI techniques to extract findings otherwise not evident with human observations or traditional analytics (Flynn & Chang, 2024).

Notably, this report considers applications in both health and environment. While much attention goes to biomedical uses, multiplexed silicon biosensors are also increasingly applied in environmental monitoring. Besides human biomarkers, their analytes can include pollutants, toxins, or pathogens in air, water, or soil (Jiménez-Rodríguez et al., 2022). Thus, two domains are examined: health (e.g. patient diagnostics, wearable health monitoring) and environment (e.g. water quality sensors, pollution detectors). The objective is to evaluate how linking silicon biosensors with Big Data infrastructure and AI analytics can enable new capabilities for real-time health and environmental data collection and intelligence. We also examine the challenges—technical, regulatory, and ethical—of deploying such systems at scale. Ultimately, we forecast a future where networks of biosensors continuously feed data into AI-driven platforms, enabling early detection of disease outbreaks,

personalized treatment adjustments, and environmental hazard warnings. The benefits of connecting biosensors to AI and Big Data include enhanced sensitivity (AI algorithms allows detection of subtle signal patterns beyond human perception), predictive modelling (forecasting health events or pollution trends from multivariate data), and automated decision-support (triggering alerts or interventions based on complex sensor-derived indicators) (Gavrilaş et al., 2022; Xi et al., 2021). For example, AI can analyse a multiplex sensor's multi-analyte output to diagnose a disease more accurately than any single-marker test, or aggregate data from hundreds of environmental sensors to predict an impending air quality issue (Gavrilaş et al., 2022; Xi et al., 2021).

At the same time, integrating biosensors with Big Data and AI raises important considerations around data governance, privacy, and ethics—especially in the health context. It is vital that these technologies are developed and deployed in a manner that respects personal data protections and maintains public trust (Bouderhem, 2024). This report provides an overview of the state-of-the-art, highlighting both the opportunities and the practical or ethical restraints that stakeholders (scientists, clinicians, policymakers, and the public) must navigate. Section 2 surveys the literature and existing technologies, detailing the role of multiplexed silicon biosensors contributing to AI-driven analytics, their mechanisms for acquiring health and environmental data, and the applicability of data protection and regulatory frameworks governing their use. Subsequently, we examine the integration and analysis of the resulting Big Data, the AI methodologies employed, representative use cases, and challenges to global-scale implementation. Ethical and legal considerations—particularly those related to informed consent, algorithmic fairness, explainability, and trust—are interwoven throughout as guiding principles and foreseeing recommendations.

2. Literature Survey

This section reviews current technologies and knowledge relevant to multiplexed silicon biosensors, their data, and AI-driven analysis, providing context for forecasting their future use. It addresses recent advancements in digital POC diagnostics, the integration of biosensors with health and environmental data streams, and the associated regulatory and ethical landscape, with emphasis on the European context. It also surveys methods of data integration, Big Data analytics techniques, and AI algorithms in use, before examining applications and broader challenges.

2.1. Significance, Forecasting & Modelling

2.1.1. Advancements in Point-of-Care Diagnostics

In recent years, point-of-care testing (POCT) has expanded from simple single-analyte tests (e.g., glucose test strips, lateral-flow pregnancy tests) toward more complex, digital, and automated systems. Traditional POCT devices were typically limited to single-marker and analog output, often delivering binary results (e.g. a coloured line on a test strip). In contrast, modern POCT technologies encompass microfluidic lab-on-a-chip systems, portable PCR analysers, smartphone-integrated diagnostic tools, and wearable biosensors capable of continuous physiological monitoring. A significant advancement is multiplexed POCT (xPOCT), which enables the simultaneous detection of multiple analytes from a single specimen at the point of care. Dincer et al., (2017) define xPOCT as the simultaneous onsite detection of different substances from a single sample, noting its increasing importance for clinical diagnostics. Dincer and colleagues (2017) review various multiplexing technologies—including bead-based assays and microarray chips utilizing optical or electrochemical detection—and conclude that the contemporary xPOCT systems are capable of simultaneously measuring approximately 10 analytes, a capacity deemed sufficient for numerous applications.

The push toward multiplexing arises from the recognition that accurate diagnoses of complex conditions—such as sepsis or stroke—often necessitates the measurement of multiple biomarkers, as no single marker provides definitive diagnostic value. Consequently, POC platforms are increasingly designed to combine high analytical performance with user-friendly operation, enabling multi-analyte testing in a single device accessible to non-specialist users. This evolution is frequently realized through miniaturized silicon biosensor arrays, where each sensor element is functionalized to detect a distinct analyte. Multiplexed silicon biosensors represent a next-generation solution in POC diagnostics, overcoming the limitations of earlier rapid tests, which commonly lack the sensitivity, quantitative precision, or multiplexing capacity of centralized laboratory assays (Laskey, 2022). By contrast, silicon-based sensor chips can incorporate sophisticated signal amplification and processing, yielding quantitative results for multiple targets with laboratory-like precision.

The impact of this technological advancement is reflected in market forecasts: the global POC diagnostics market is expected to more than double in the mid-2020s, rising from an estimated \$4.8 billion in 2019 to \$10.1 billion by 2025 (representing a compound annual growth rate of ~13%).

Withing this trend, the biosensor-based device segment—particularly platforms utilizing microfluidic chips—is forecasted to grow at ~7.7% annually (Laskey, 2022). Industry analysts attribute this growth to the advantages of integrated biosensor platforms—miniaturization, high sensitivity, robustness, multiplexing capability, and cost-efficient scalability (Laskey, 2022). In short, multiplexed biosensors are poised to make diagnostics faster, more affordable, and broadly accessible, an advancement with substantial implications for global healthcare delivery.

2.1.2. Contributions to AI & Predictive Modelling

Another key dimension of the importance of multiplexed biosensors lies in their integration with the digital data ecosystem and their enabling role in A-driven medicine. In contrast to traditional analog tests, which are interpreted visually, digital biosensors generate machine-readable data streams that can be automatically captured, transmitted, and analysed by computational systems. The multi-parametric output of multiplexed biosensors provides rich, high-dimensional datasets ideal for input into machine learning (ML) models for pattern recognition, anomaly detection, and predictive diagnostics.

For instance, simultaneous measurements of several inflammatory markers from a multiplex biosensor can be processed by a trained AI model to predict or classify disease states with greater accuracy than conventional single-marker threshold approached. Han et al. (2025) emphasize that incorporating ML algorithms into POCT platforms significantly enhances diagnostic performance. An illustrative example is AI-powered image analysis of lateral-flow assays, which can detect faint positive lines that might be missed by human interpretation (Han et al., 2025).

Notably, Han and colleagues point out current limitations in the field: many existing POCT platforms have limited multiplexing and lack advanced analytical algorithms to interpret complex multivariable patterns, which restricts their use for co-infections or multi-biomarker panels (Han et al., 2025). This constrains their effectiveness in scenarios requiring simultaneous detection of co-infections or analysis of biomarker panels. These observations reinforce the interdependence of multiplexing and AI—realizing the full potential of multiplex diagnostics requires sophisticated algorithms capable of extracting insights from multivariate patterns. Conversely, AI-based predictive modelling in healthcare benefits from rich, multi-feature input data, which multiplexed biosensors inherently provide.

During the COVID-19 pandemic, this synergy was exemplified by ML models developed to interpret outputs from multiplex biosensors—such as combined results for multiple viral gene targets or antibodies—to infer a patient’s infection or immunity status (Beisenova et al., 2023). These models could discern complex relationships (e.g., ratios between biomarkers) that would not be readily apparent to human observers.

In the context of environmental health, data from distributed networks of multiplex sensors (e.g., measuring various pollutants) can similarly feed into predictive models forecasting air quality indices or contamination spread (Gavrilaş et al., 2022). This provides communities with early-warning systems

for emerging environmental risks, further illustrating the transformative role of multiplexed biosensors in data-intensive, AI-supported decision-making systems.

2.1.3. Bio-Inspired Manufacturing Innovations

A significant innovation advancing the development of multiplexed biosensors is the use of bio-inspired assembly techniques in their manufacturing. These techniques leverage biological molecules or design principles to assemble sensor components at the micro- or nano scale with high special precision. A leading example is the current *BioAssembler* project, which aims to integrate bio-inspired assembly into semiconductor fabrication to enable scalable production of silicon-based, label-free multiplex biosensors (Puumala et al., 2023).

One of the longstanding challenges in biosensor fabrication is that, while the production of electronic chip components is highly scalable through established semiconductor processes, the addition of biological sensing elements—such as immobilized antibodies or aptamers arranged in specific surface patterns—remains complex and difficult to scale. The *BioAssembler* project addresses this bottleneck by proposing “bio-intelligent” manufacturing strategies that fuse silicon microfabrication with biologically guided assembly. Techniques such as DNA origami and protein-directed self-assembly are being explored to position biosensing nanostructures with molecular-level accuracy onto sensor chips.

If successfully implemented, this approach could yield a new class of MEMS-based biosensors that are simultaneously mass-producible and customized with multiple bioreceptors, allowing for multiplexed detection from minimal sample volumes. According to project partners such as Biomensio, this could enable the creation of disposable diagnostic chips capable of detecting multiple health or environmental biomarkers from a single drop of fluid (Biomensio, 2022).

The potential applications of wafer-scale, biofunctionalized sensor arrays are broad and impactful—ranging from personal health monitoring and food safety diagnostics to drug screening and environmental sensing (Biomensio, 2022). Crucially, the project also incorporates a sustainability perspective by engaging in social science research to evaluate public acceptance, ethical considerations, and economic implications, promoting a responsible and inclusive transformation of the European biosensor manufacturing landscape (Puumala et al., 2023).

In summary, by addressing the critical challenge of scalable biofunctionalization, bio-inspired assembly techniques pave the way for the widespread deployment of multiplexed biosensors. This, in turn, supports richer data generation for AI-driven analytics and modelling, reinforcing the broader integration of digital diagnostics into health and environmental systems.

3.1. Health Data Collection

Silicon-based multiplex biosensors are capable of capturing a broad spectrum of health-related data, encompassing both molecular biomarkers and physiological parameters. The specific types of data acquired depend on the sensor’s design. Some configurations are optimized for biochemical analytes—such as glucose, lactate, cholesterol, hormones, cytokines, nucleic acids (DNA/RNA

sequences), and microbial antigens (e.g., from viruses or bacteria)—while others focus on physical parameters like temperature, heart rhythm (via electrical signals), or blood oxygen saturation (via optical sensing). In general, biosensors may target ions, small molecules, nucleic acids, proteins, pathogens, or even whole cells (Jiménez-Rodríguez et al., 2022). For example, a single multiplexed lab-on-a-chip device could simultaneously quantify electrolyte concentrations (e.g., sodium, potassium), metabolic markers (e.g., glucose, lactate), and inflammatory indicators such as C-reactive protein. These outputs are typically quantitative, delivering precise concentration values, rather than qualitative binary results. Data collection can be episodic, such as with cartridge-based tests used on demand, or continuous, as in wearable or implantable biosensors designed for ongoing monitoring.

One advantage of digital multiplex biosensors is their compatibility with Electronic Health Records (EHRs) and broader clinical data systems. By combining biosensor output with patient-specific clinical data—such as medical history, medications, and treatment reports—clinicians and data scientists can gain a more comprehensive understanding of health status. For instance, real-time blood pressure or blood glucose can be automatically appended to patient’s longitudinal EHR timeline, providing context for concurrent medications or reported symptoms. Moreover, biosensor data can be compared against conventional laboratory results in the EHR to validate accuracy and track physiological trends over time.

Continuous monitoring is particularly impactful. Wearables and at-home biosensors enable frequent sampling, for example, continuous glucose monitor readings every few minutes or a smart patch measuring vital signs around the clock. This yields longitudinal datasets rather than single snapshots; clinical benefits of such continuous data are well documented. Smith et al. (2023) report that chronic disease patients using wearable biosensors to monitor multiple daily metrics are better able to adjust their behaviour and medication use in real time, reducing the incidence of acute exacerbations and hospitalizations. Such self-monitoring has been linked to improved disease management and cost savings. For healthcare professionals, integrated sensor-EHR datasets enable earlier detection of deteriorating conditions, while researchers can analyse population-level patterns, such as how real-world blood pressure variability correlates with medication adherence and effectiveness.

To realize these benefits at scale, biosensor data must integrate seamlessly and securely with clinical IT infrastructure, including hospital EHRs and regional health information exchange. Standards such as HL7 FHIR (Fast Healthcare Interoperability Resources) are increasingly used to ensure structured, consistent data exchange. For example, a smart glucometer may transmit data via Bluetooth to a mobile app, which then uploads the data to the patient’s clinic record using a FHIR-compliant API. Some healthcare providers have begun integrating patient-generated health data into their systems, although broader adoption is still limited. Surveys indicate only around 10% of physicians routinely incorporate wearable sensor data into EHRs, due to workflow burden and data quality concerns. Nonetheless, as the technical standards mature and confidence in the reliability of sensor data increases, integration is expected to become more routine.

The European Health Data Space (EHDS) initiative further envisions a framework that allows individuals to securely share data (including sensor data) across EU member states (Joint Action Towards the European Health Data Space TEHDAS, 2023). A practical example of such integration in practice is emerging in remote patient monitoring programs: patients with chronic conditions (say, heart failure) may be given a kit with biosensors (e.g. a connected blood pressure cuff, weight scale, or a blood test device) that automatically transmits daily measurements to the clinic’s system. Nurses are alerted if values cross certain thresholds, enabling early intervention. This integration of multiplex sensor data into routine care has been shown to improve outcomes in pilot studies and is being scaled up in many health systems.

3.2. Health Data Protection & Ethical Considerations

Collecting and utilizing data from multiplexed biosensors involves processing sensitive personal health information, subject to multiple regulatory and ethical frameworks. In the European Union, the primary legal instrument is the GDPR, which designates health data as a special category requiring enhanced protection. GDPR enforces principles such as data minimization, purpose limitation, and a lawful basis for data processing (e.g., explicit consent or processing in the public interest) (European Parliament & Council of the European Union, 2016). For instance, if a wearable biosensor vendor collects heart rate or blood analyte data, it must obtain informed consent (or invoke another valid basis) and transparently communicate data use to the user.

GDPR grants individuals rights over their personal data—access, rectification, erasure, and portability—which apply equally to biosensor data as to EHRs. Anonymized data are excluded from GDPR scope, making anonymization or pseudonymization standard practice for secondary uses such as AI model training. Flynn & Chang (2024) emphasize the need for anonymization and ethical governance in secondary processing of biosensor datasets.

Beyond GDPR, additional EU regulations are emerging (**Figure 1**). The EU AI Act, currently under implementation, introduces a risk-based regulatory framework for artificial intelligence. Most healthcare-related AI systems, including those interpreting biosensor data, will be considered “high-risk” and subject to requirements for transparency, accuracy, risk management, technical documentation, and human oversight. For example, an AI-enabled biosensor controlling insulin dosage would need to be traceable and explainable to users and regulators. The Act prohibits certain applications (e.g., social scoring) and mandates fairness, including addressing algorithmic bias. Boudierhem (2024) notes its potential as a model for global biosensor regulation.

Another key initiative is the EHDS, which aims to enable secure use of health data for care (primary use) and research or innovation (secondary use). EHDS will require safeguards such as pseudonymization and secure processing environments, preventing raw personal data from being downloaded or misused. Researchers may only access aggregated data or query person-level data through approved, controlled interfaces. This clarifies how biosensor datasets can be legally pooled across EU member states (European Commission, 2022).

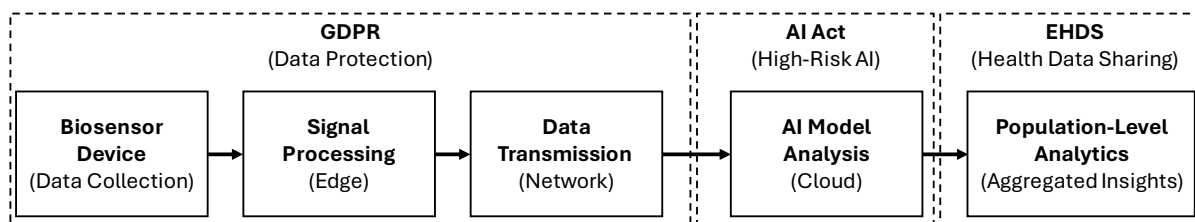


Figure 1. Key EU regulatory checkpoints mapped onto the biosensor data lifecycle. GDPR imposes strict requirements on personal data from the moment of collection and throughout processing, mandating privacy protections and lawful use of sensitive biosensor-derived information. As the data pipeline incorporates AI-driven analysis, the AI Act comes into play – high-risk AI systems (e.g. diagnostic algorithms using biosensor data) must meet safety, transparency, and oversight criteria before deployment. For secondary use and broader data sharing, the proposed EHDS framework governs the secure exchange of health data across the EU, enabling research and public health analytics while ensuring compliance with data protection standards. These regulations collectively ensure that data handling at each stage, from initial sensing to AI processing and data sharing, adheres to EU privacy, safety, and ethical guidelines.

Securing biosensor data requires both technical and organizational safeguards. Encryption should be applied on-device, during transmission, and at rest. Many biosensors already implement these protections (e.g., encrypted Bluetooth and HTTPS upload). Advanced techniques include blockchain, for tamper-evident access logging, and federated learning (FL), which trains AI models without transferring raw data. In FL, model training occurs locally, and only model parameters are aggregated centrally (**Figure 2**). This approach complies with privacy standards while enabling cross-institutional learning. FL can be further enhanced with differential privacy to mask individual contributions (Han et al., 2025; Kalodanis et al., 2025). These measures are critical for maintaining user trust and preventing reputational damage from data breaches.

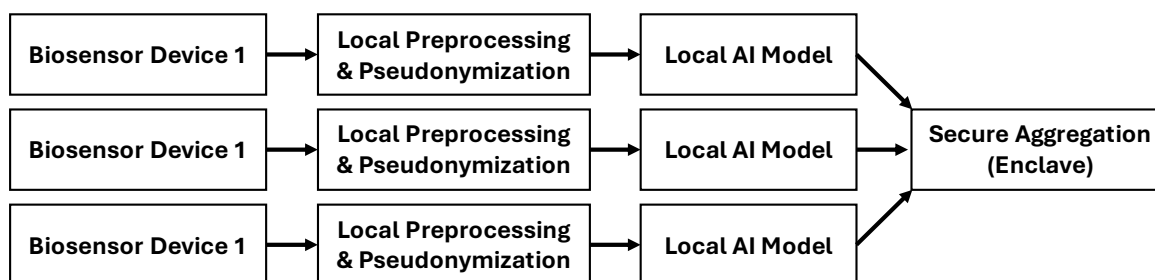


Figure 2. Privacy-preserving techniques applied across the biosensor data pipeline. Pseudonymization is performed at the point of data capture or immediately after, replacing or removing personal identifiers from biosensor data to reduce re-identification risk. Instead of sending raw sensitive data to the cloud, federated learning keeps data on multiple edge devices and only exchanges model parameters – this collaborative training approach addresses privacy by ensuring that individuals’ data never leaves their device. At the aggregation and analysis stage, secure enclaves (trusted execution environments) can be employed on cloud servers to process data or model updates in encrypted form, guaranteeing that even the cloud provider cannot directly access the raw personal data. Layering these techniques – pseudonymizing data, training AI models locally, and using encrypted computation – provides defense-in-depth for sensitive health information in AI-enabled biosensor networks.

Cross-border data sharing introduces additional legal complexities. Within the EU, data can move freely under GDPR, but transfers outside the EU require equivalent protection, often managed through Standard Contractual Clauses. Re-identification risks increase when biosensor data are combined with external sources (e.g., patient registries), even if initially anonymized. Controllers must assess these risks and may apply techniques such as k-anonymity, outlier suppression, or summary statistics to mitigate re-identification (European Commission, 2019).

The EU's AI governance also emphasizes explainability and user rights. High-risk AI systems must provide intelligible explanations for outputs. In biosensor applications, this may involve displaying threshold-based triggers or clinician-accessible feature attribution. Explainable AI (XAI) methods—such as SHAP values, saliency maps, or surrogate models—support this requirement and align with ethical AI principles, including transparency, fairness, and human oversight (Sullivan, 2023).

Development and use of AI models in medical settings is a further consideration. Early-stage ethical assessment of AI tools is encouraged, although their long-term validity remains uncertain as technologies mature. Regulatory sandboxes have been highlighted as a mechanism for testing novel services and legal interpretations under controlled conditions (DigiFinland Oy, 2024). These simulated environments allow safe experimentation with new AI models without undermining institutional compliance. As outlined in Finland's public sector guide to responsible AI use (Suomi.fi, n.d.), success depends on active participation across service providers, regulators, and technology developers. Promising use cases include automated patient documentation systems—sometimes referred to as “nurse whisperers”—and AI-assisted diagnostic tools, whose feasibility and impact should be jointly evaluated.

Additional ethical concerns include informed consent, profiling, and secondary use. Individuals should be informed not only of the primary use of their biosensor data but also of potential secondary uses, such as research. GDPR requires specific consent for health data unless another legal basis applies. There is growing support for dynamic consent models, allowing participants to manage preferences via digital tools. This aligns with EU efforts to support data autonomy, such as the "My Rights Over My Health Data" portal (European Commission, Directorate-General for Health and Food Safety, n.d.).

Profiling (the use of data to infer health conditions or behaviours) presents ethical risks, particularly if used in discriminatory contexts (e.g., by insurers or employers). Under GDPR Article 22, individuals have rights regarding automated decision-making and profiling with significant effects. Even with consent, such use cases require robust oversight and safeguards to avoid misuse and erosion of public trust.

In conclusion, the deployment of AI-driven biosensor systems necessitates strong legal and ethical foundations. The EU model—centered on privacy by design, user rights, and accountability—offers a comprehensive approach. However, operational challenges remain: balancing transparency with usability, enabling data utility while preserving privacy, and streamlining compliance processes to avoid stifling research. As noted by , Brück et al., (2024), overly burdensome regulation can hinder

innovation. Going forward, sustaining public trust will require a combination of technical safeguards, clear communication, and inclusive governance involving data subjects in decision-making. Ethical use of biosensor data is as vital as the underlying technological innovation (Sullivan, 2023).

3.3. Environmental Measurement Data

Beyond human health, multiplexed silicon biosensors can be deployed to collect environmental data that are crucial for ecological monitoring and public health protection. These sensors can be designed to detect a variety of environmental parameters and contaminants. For example, biosensors have been created to monitor water quality by concurrently detecting multiple pollutants such as heavy metals (e.g. lead, mercury), nutrients (nitrate, phosphate), and organic toxins (pesticides, industrial chemicals) in a single device (Gavrilaş et al., 2022). In air quality monitoring, silicon-based sensor arrays might measure particulate matter levels along with gaseous pollutants like volatile organic compounds, carbon monoxide, nitrogen dioxide, and ozone all at once. Some biosensors even use biorecognition elements to detect biological agents in the environment, such as specific bacteria or viruses in water or soil samples (Jiménez-Rodríguez et al., 2022). For instance, a multiplex biosensor could screen a water sample for several pathogens (e.g. *E. coli*, *Salmonella*, *Giardia*) concurrently, providing an overall microbial safety readout. Thus, the types of environmental data measurable are broad: chemical concentrations, presence of biological pathogens, and even physical parameters (if the sensor integrates, say, temperature or pH sensing alongside biochemical detection).

Such environmental data are invaluable for Big Data analytics in environmental science. Environmental sensors are often deployed in distributed networks (for example, dozens of sensors across a geographic area) producing continuous streams of readings. When aggregated, these streams form large datasets ripe for analysis. Big Data methods can uncover spatiotemporal patterns – for example, ML might analyse data from hundreds of multiplexed air sensors in a city to identify pollution hotspots and peak times, correlate those patterns with traffic or industrial activity, and predict future pollution episodes. Indeed, analytics platforms for “smart cities” are being developed to integrate multi-sensor inputs (weather, air quality, noise, etc.) and provide city managers with real-time situational awareness and forecasting. Multiplex biosensors enhance this by providing richer data per sensor node (each sensor capturing multiple pollutants or factors rather than just one).

Environmental monitoring use cases span several areas:

- **Air pollution monitoring:** Cities can install sensor nodes on lamp posts or buildings that continuously measure multiple pollutants. Combined with meteorological data, these inputs can feed AI models that predict poor air quality days, allowing authorities to issue health advisories or implement traffic restrictions proactively. Over the long term, the collected data help evaluate policy interventions (e.g. determining whether establishing a low-emission zone reduced certain pollutant levels).
- **Water quality and safety:** In water treatment plants or distribution systems, multiplex biosensors can continuously check for a panel of contaminants (for example, heavy metals,

nitrites, and microbial toxins). If any exceed safe thresholds, the system can trigger alarms far faster than periodic manual lab tests would. In the field, portable multiplex sensors can be used during environmental emergencies (like floods or chemical spills) to rapidly assess contamination spread and identify areas of concern.

- **Marine and coastal monitoring:** A case study from recent research demonstrated that a portable flow-based multiplex device could be used for continuous in situ bio-detection of seawater contaminants (Salvador et al., 2022). In that study, the device simultaneously tracked multiple pollutants in seawater, illustrating the potential for continuous monitoring in marine environments (e.g. detecting algal toxins, oil residues, and heavy metals in near real time).
- **Agriculture and food:** Biosensors are being applied to measure soil health parameters; for instance, a multiplex sensor stake inserted in soil might track moisture, pH, and nutrient levels (nitrogen, phosphorus, potassium) to guide precision agriculture. Farmers could get a fine-grained map of soil conditions and apply water or fertilizer only as needed and where needed. Additionally, sensors in greenhouses could monitor multiple climate and soil factors (CO₂, humidity, nutrient content in hydroponic solution) to optimize plant growth. In food safety, multiplex biosensors could test for multiple pathogens or toxins in food products on processing lines, preventing contaminated batches from reaching consumers by providing rapid on-site screening.
- **Biodiversity and ecosystem monitoring:** Environmental DNA (eDNA) biosensors are being developed to detect traces of DNA from various species in water or soil, which can indicate biodiversity levels or the presence of invasive species. A single eDNA biosensor sample could, for example, simultaneously check for DNA of dozens of fish or amphibian species in a river, helping conservationists monitor species distribution without having to physically observe the organisms.

Scaled deployments of multiplex sensors generate multi-dimensional, time-series data that can be linked with public health outcomes. For example, urban health initiatives use environmental sensor networks in combination with hospital admission data to model the health impacts of pollution exposure (Joint Action Towards the European Health Data Space TEHDAS, 2023). Vehicle-mounted sensors can map pollution by street, which, when cross-referenced with health statistics, allows AI to predict outcomes such as asthma spikes based on pollution mixtures. These efforts typically rely on community-level aggregated health data to preserve privacy, with public health institutions serving as neutral data stewards. The upcoming EHDS initiative aims to support such cross-domain analytics by offering clear frameworks for health-environment data integration across regions (European Commission, 2022).

Overall, multiplexed environmental biosensors combined with Big Data analytics enable a shift from reactive to proactive environmental management. Patterns of pollution or contamination that were previously difficult to discern can now be detected early, and interventions can be evaluated with data driven evidence. These systems, however, raise analogous issues of governance and trust as in health:

data quality (sensor drift or fouling must be managed), data interoperability (standard formats for environmental data reporting), and ethical use (ensuring, for example, that sensor data in agriculture isn't misused to disadvantage small farmers). While environmental data generally do not involve personal privacy to the same extent as health data, there can still be sensitivities. Neutral data governance structures, such as cooperatives or public repositories, may be needed to ensure fair use and sharing of environmental biosensor data. In agriculture, for example, the concept of a common European agricultural data space is being explored to help farmers share and benefit from data collectively without fearing exploitation (Joint Action Towards the European Health Data Space TEHDAS, 2023).

3.4. Data Integration

To fully realize the value of multiplexed biosensors, their data must be integrated with other relevant datasets. This integration can occur at multiple levels, including sensor-level fusion (combining streams from multiple biosensors), clinical integration (linking biosensor outputs with electronic health records, genomic profiles, or medical imaging), and contextual integration (combining health sensor data with external variables such as environmental exposures or behavioural patterns).

At the technical level, integration is commonly implemented through data warehouses or data lakes, which serve as centralized repositories for raw or lightly processed data from various sources (**Figure 3**). Increasingly, real-time data exchange is enabled via application programming interfaces (APIs). For example, a hospital information system might retrieve data from a patient's wearable device through an API provided by the device manufacturer.

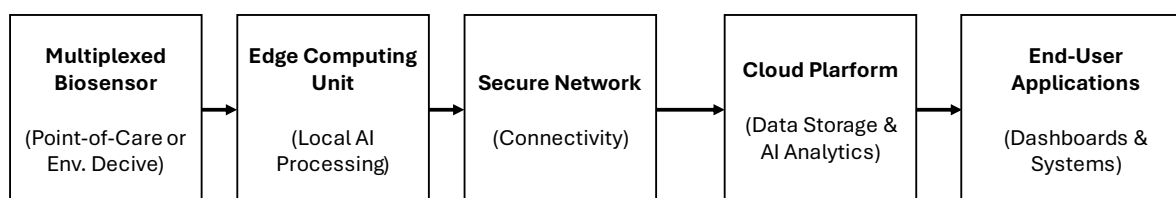


Figure 3. Modular architecture for deploying AI-enabled multiplexed biosensors in point-of-care or environmental settings. The design follows a layered IoT architecture. At the field level, multiplexed biosensor units (each incorporating the sensor and on-site electronics) collect data and may perform initial edge AI processing. A secure connectivity layer links these devices to the cloud via gateways or wireless networks (e.g. BLE, 5G), ensuring encrypted data transmission. The cloud platform ingests and stores the biosensor data and hosts AI analytic services to further process and analyse large-scale inputs. Finally, an application layer provides end-user interfaces—such as clinician dashboards, public health monitoring portals, or environmental management systems—that allow stakeholders to access insights and integrate them into decision-making. This modular setup supports scalability and flexibility, enabling biosensor deployments to feed into centralized analytics while respecting privacy and regulatory constraints.

Interoperability is essential for scalable integration. In the health sector, HL7 FHIR provides standardized data structures for representing common clinical entities, such as observations and medications. A biosensor reading can be represented as a FHIR Observation resource, containing fields for value, unit, timestamp, and patient reference, and can be seamlessly ingested into any FHIR-

compliant electronic health record system. In the environmental domain, similar standards are emerging, including MQTT (Message Queuing Telemetry Transport) and the OGC SensorThings API, which define formats and communication protocols for Internet of Things (IoT) sensor data. The adoption of common standards ensures compatibility between systems and facilitates cross-domain data integration.

Privacy protection is a critical concern when integrating personal biosensor data with other datasets. Linking data at the individual level must be done in a way that prevents re-identification. Techniques such as pseudonymization—where real identifiers are replaced with study-specific codes—and privacy-preserving record linkage—using cryptographic hashes and trusted intermediaries—allow datasets to be linked without exposing personal identities. In Finland, for example, the national data permit authority FinData enables secure linkage by acting as a neutral intermediary. Researchers submit data requests, and FinData handles data extraction, pseudonymization, and provisioning within a secure analysis environment, without disclosing full identifiers to any party.

Integrating biosensor data with other modalities yields multimodal datasets that offer deeper analytical insights than any single data source alone. For example, combining clinical data (such as diagnoses or medication history) with biosensor readings (vital signs, biochemical markers) and environmental data (like air quality or temperature) can improve the accuracy of predictive models, such as those used to forecast asthma attacks. A recent study by Prissel et al., (2023) demonstrated the value of cross-domain integration by linking groundwater contamination data with pediatric health records, uncovering associations between nitrate exposure and adverse health outcomes that would not be evident from health or environmental data alone.

From a Big Data perspective, integration addresses the variety dimension—joining time-series sensor data with relational clinical data, geospatial records, and other sources. Temporal and spatial alignment is crucial. For instance, a biosensor's heart rate reading must be matched with the closest blood test result in time, or a pollution measurement must be linked to patients living in the affected area. Rich metadata are essential for this alignment: each sensor reading should include time, location (for environmental sensors), patient identifiers (when appropriate), and technical context about how the data were captured.

The EHDS initiative supports this kind of integration by promoting standardized data access formats and federated architectures (European Commission, 2022). Rather than centralizing data across regions, the EHDS envisions a networked model in which researchers can securely query distributed datasets without moving or exposing raw data. This approach respects data sovereignty while enabling cross-border research collaboration.

Accessing non-health datasets—such as meteorological, traffic, or socioeconomic data—often requires additional permissions. While some sources (e.g., European Copernicus satellite data, public weather station data) are open-access, others (e.g., detailed mobility patterns, private fitness app data) may be proprietary. Integration projects must therefore establish formal data-sharing

agreements that define permitted uses and ensure that when data are combined, the resulting dataset is governed by privacy protections appropriate to its most sensitive component. Often, integration occurs within secure computing environments that permit data linkage and analysis without enabling full data export or re-identification. This model is already in use across many EU countries via national health data permit authorities and will likely be reinforced under the EHDS regulatory framework.

Effective data integration is foundational to extracting value from multiplexed biosensors. It enables context-rich analysis, allowing biosensor data to be interpreted in relation to medical history, environmental conditions, and other external factors. A biosensor reading in isolation may signal a physiological abnormality, but when combined with clinical and contextual data, it can support more accurate diagnoses, trigger personalized interventions, or inform public health alerts. Achieving this requires robust data interoperability, secure infrastructure, and strict adherence to privacy principles.

Current efforts to standardize APIs and data formats, along with investments in trusted data environments, are helping lay the foundation for such integration. As interoperability improves, new applications are emerging—such as telehealth platforms that merge real-time biosensor data with EHRs and environmental feeds to support clinical decision-making. In remote or underserved areas, integrated systems may allow specialists to monitor patient data from a distance, extending expert care to rural clinics. The long-term vision is for biosensor data to flow securely and appropriately across domains—reaching clinicians, researchers, and public health authorities as needed—while upholding patients’ rights, preserving confidentiality, and ensuring that each actor accesses only what is necessary for their role.

3.5. Big Data Analysis

The proliferation of multiplex biosensors and integrated data streams introduces both significant opportunities and complex challenges associated with Big Data. This landscape is commonly described using the “4 Vs”: **Volume, Velocity, Variety, and Veracity**.

In terms of **volume**, multiplex biosensors generate large-scale datasets. Continuous monitoring from a single individual can yield thousands of data points per day, and when scaled across hundreds or thousands of sensors in a network, the cumulative data can rapidly grow to terabytes. **Velocity** refers to the real-time nature of data generation; biosensors stream data continuously from edge devices, requiring real-time ingestion and analysis pipelines rather than batch-processing approaches. **Variety** encompasses the diverse formats and types of data produced, from structured numeric time-series to categorical alerts and even unstructured modalities such as images or textual notes linked to sensor readings. **Veracity** addresses data quality, which is a non-trivial concern in biosensor data due to issues like sensor noise, calibration drift, missing data, or hardware faults.

Analysing such data involves a combination of classical statistical methods, modern ML, and large-scale data engineering. Time-series analysis remains foundational, employing techniques such as moving averages, Fourier transforms, or wavelet analysis to identify trends, periodicities, and

anomalies. For example, continuous ECG signals from wearable biosensors can be analysed to detect arrhythmias or circadian patterns using signal processing methods that interpret peak structures and waveform intervals.

At larger scales, processing biosensor data often requires distributed computing frameworks like Hadoop or Spark. These allow parallelized analysis over data clusters, which is essential when dealing with sensor data aggregated from multiple sources or geographic regions, such as EU-wide sensor networks envisioned under the EHDS.

Pattern recognition is central to biosensor analytics. Unsupervised machine learning methods, such as clustering, can uncover latent structures in the data—grouping similar pollution exposure profiles, or identifying subpopulations with shared biosensor-derived physiological patterns. For predictive tasks, supervised learning models are trained on historical biosensor data combined with known outcomes to forecast future events. For instance, a model might use heart rate, weight, blood pressure, and oxygen saturation patterns to predict heart failure exacerbations. These models often built using deep neural networks, random forests, or gradient boosting, can manage complex, high-dimensional inputs and provide predictions that go beyond simple threshold-based alarms.

Real-time analysis is especially critical in point-of-care applications. Techniques like complex event processing (CEP) enable detection of predefined patterns across multiple sensor streams, signalling clinically relevant events such as impending asthma attacks. For continuous data, online learning algorithms and streaming statistics update model parameters incrementally as new data arrive, supporting real-time anomaly detection and personalized baseline modelling.

When biosensor data are pooled from many individuals or institutions, new opportunities and challenges arise. Federated analytics, an emerging paradigm, allows analysis across distributed datasets without transferring sensitive data (**Figure 2**). Instead of centralizing raw data, analysis code is executed locally in each data repository, and only aggregated or de-identified results are shared. This approach aligns with EHDS principles and European data sovereignty, enabling large-scale analytics while preserving privacy. Effective implementation requires harmonized data structures and metadata (e.g., ensuring consistent units for glucose measurements across datasets) and robust statistical methods to combine results across sites.

Computational scalability is a critical consideration. Training complex AI models on multimodal biosensor data from tens or hundreds of thousands of individuals likely require GPU acceleration and cloud infrastructure. Efficient algorithm design and optimized resource allocation are hence essential for managing these workloads, and many cloud platforms now offer managed services for distributed analytics, albeit with technical and cost considerations.

In addition, addressing data quality and uncertainty is essential. Sensor signals can be noisy or incomplete due to device issues or environmental interference. Techniques such as Kalman filters or particle filters help extract reliable signals from noisy input. Missing data—common with wearables

due to connectivity loss or noncompliance—must be handled carefully, especially when “missingness” (e.g., *NA* values) is non-random. Imputation methods and Bayesian models capable of modelling “missingness” mechanisms help reduce bias. Data validation pipelines are also required to flag implausible values or malfunctioning sensors.

Exploratory analyses across large biosensor datasets must be statistically rigorous to avoid false discoveries. When patterns are identified without prior hypotheses, proper use of validation datasets, correction for multiple testing (e.g., Bonferroni adjustment or false discovery rate control), and cross-validation procedures are necessary to ensure findings are generalizable.

Advanced ML architectures are increasingly being explored for biosensor data. Transformer models, originally developed for natural language processing, show strong performance in modelling long-range dependencies in multivariate time-series and are outperforming traditional recurrent neural networks (RNNs) in some applications (Han et al., 2025). Convolutional neural networks (CNNs) are also applied by representing biosensor data as matrices or images (e.g., time vs. analyte), enabling the extraction of spatial and temporal features from multiplexed sensor outputs.

Interpretability remains a critical requirement, particularly in healthcare contexts. Even when using complex models, clinicians and decision-makers must understand the basis of predictions. In many regulatory contexts, especially in medicine, “black box” models are viewed cautiously unless they can provide at least some rationale for decisions (European Commission, 2019). This drives the use of explainable AI (XAI) techniques, such as SHAP values, attention maps, or surrogate rule-based models, to provide post-hoc explanations. In some cases, simpler models like decision trees are preferred for their transparency. Interpretability facilitates trust, regulatory compliance, and actionable insights—for instance, identifying combinations of environmental and physiological factors that precede adverse events such as asthma exacerbations.

As data cannot always be centralized due to privacy or regulatory constraints, federated and distributed analytics offer a path forward. These approaches, already proven in sectors like finance and being piloted in healthcare, allow model training and statistical computation across siloed datasets. For example, intensive care unit (ICU) monitoring models have been trained via federated learning across hospitals, with each site retaining control over its data and contributing only model updates. Similar strategies can apply to querying and descriptive analytics using map-reduce logic, enabling regional insights without compromising individual privacy.

In conclusion, Big Data analytics techniques are essential for converting high-**volume**, high-**velocity**, and high-**variety** biosensor data into actionable insights. By leveraging advanced statistical methods, scalable infrastructures, and privacy-preserving architectures, researchers and clinicians can detect early warning signals, personalize interventions, and better understand complex health-environment interactions. As data availability continues to grow, the main bottleneck will increasingly be our ability to interpret and operationalize the information in ways that are clinically and societally relevant. With

careful methodological design, these datasets offer unprecedented potential to advance personalized and population health.

3.6. Artificial Intelligence (AI)

AI is playing an increasingly pivotal role in extracting value from multiplexed biosensor data, enhancing everything from raw signal interpretation to real-time clinical decision-support. A range of AI techniques is applied across this pipeline, spanning traditional machine learning algorithms and state-of-the-art deep learning models.

A foundational use of AI in biosensor systems is signal processing. Many biosensors produce raw physical signals—such as electrical impedance changes or optical resonance shifts—that require transformation into clinically meaningful metrics (**Figure 4**). AI models, particularly neural networks, can be trained to map these raw outputs to quantitative analyte concentrations more accurately than traditional linear calibration methods. This is particularly beneficial when sensor responses are nonlinear or influenced by confounding environmental conditions. Flynn & Chang (2024) emphasize that AI can be integrated across the full biosensor lifecycle—from sensor design and calibration to data interpretation—enabling more accurate, adaptive, and robust sensing platforms. For example, multi-analyte biosensors can benefit from AI-driven calibration algorithms that account for cross-sensitivities and learn compensation strategies for complex signal interdependencies.

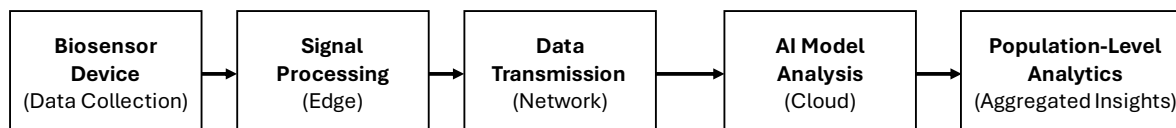


Figure 4. Hierarchical end-to-end data pipeline from biosensor readings to aggregated insights. Multiplexed silicon biosensors at the point-of-care or in the environment collect raw biological or chemical signals, which are digitized and pre-processed at the edge. The processed data is then transmitted over secure networks to an AI integration layer for analysis. Machine learning models evaluate the incoming biosensor data to detect patterns or anomalies, and the results are aggregated into population-level health or environmental analytics. This flow enables real-time monitoring and extraction of epidemiological or environmental trends from distributed sensor inputs.

For POC diagnostics, AI often needs to be deployed locally on the device or on a connected smartphone—referred to as edge computing. In these contexts, rapid inference is critical, and cloud connectivity may be unreliable or unavailable. This has driven the adoption of TinyML—machine learning models optimized for execution on microcontrollers or low-power chips. Examples include glucometers that use embedded AI models to correct for external influences such as ambient temperature or patient hematocrit, or wearable heart monitors that perform real-time arrhythmia detection using compact neural networks. Edge AI systems must be optimized for memory and energy efficiency, often through model compression techniques like quantization or knowledge distillation. Despite hardware constraints, edge AI enables biosensor platforms to operate autonomously, delivering fast and reliable outputs with minimal latency.

As biosensors are deployed more widely, the resulting availability of large, labelled datasets further enhances AI performance. When outcomes from thousands of multiplex diagnostic tests are recorded, supervised learning models can be trained to detect complex biomarker signatures associated with specific diseases. These models can significantly outperform simple threshold-based rules by capturing nonlinear interactions across multiple analytes. Convolutional Neural Networks (CNNs) are effective when biosensor outputs have spatial structures, such as when interpreting patterned responses in microfluidic devices or assay arrays. For temporal data streams—common in wearable monitoring—Recurrent Neural Networks (RNNs) and Transformer architectures are increasingly used. Transformers, in particular, have shown strong performance in time-series tasks due to their ability to model long-range dependencies and parallelize learning over extended input sequences (Han et al., 2025). For instance, a Transformer could process continuous data streams from multiple vital signs to predict clinical deterioration or disease flare-ups several days in advance.

Overall, AI significantly extends the capabilities of multiplexed biosensors, improving their accuracy, autonomy, and clinical utility. As both model sophistication and deployment infrastructure continue to evolve, AI will remain central to unlocking the full diagnostic and predictive potential of next generation biosensing systems.

Some key AI algorithms and approaches used in biosensor analytics include:

- **Decision Trees and Random Forests:** Commonly used for classification tasks such as predicting disease presence from biosensor outputs. These models capture nonlinear relationships and feature interactions effectively and offer built-in interpretability through feature importance metrics. For example, a random forest might identify which subset of biomarkers in a multiplexed panel contributes most significantly to a diagnostic decision.
- **Support Vector Machines (SVMs):** Applied particularly in scenarios involving complex signal patterns, such as in electronic “nose” or “tongue” sensors. SVMs perform well in high-dimensional spaces and can model nonlinear boundaries using kernel functions. For instance, they can classify chemical sensor array outputs to detect specific compounds or contaminants.
- **Neural Networks (Deep Learning):** Various architectures are used depending on the data structure. Feed-forward networks are suited for regression tasks (e.g., mapping raw signals to analyte concentrations), CNNs are employed to extract features from image-like or spectral sensor data, RNNs (such as LSTMs) are effective for analysing temporal sequences from continuous monitoring, and Transformer models are increasingly used for multivariate time-series due to their ability to capture long-range dependencies. For example, CNNs may be used to analyse the spectral signature of a sensor response, while LSTMs can monitor evolving patient vitals to predict critical events.
- **Clustering and Dimensionality Reduction:** Unsupervised methods such as k-means clustering help reveal patterns or patient subgroups within high-dimensional biomarker data. Dimensionality reduction techniques like PCA, t-SNE, or UMAP are used to simplify complex

datasets and facilitate visualization or preprocessing. These approaches can identify natural groupings in biosensor profiles or reduce input complexity before modelling.

- **Transfer Learning:** Useful when adapting models trained on large biosensor datasets to new, smaller datasets. A pre-trained model can be fine-tuned on a new device or population, improving performance without requiring extensive retraining. For instance, a cardiac anomaly detection model trained on adult data could be adapted to pediatric data using transfer learning.
- **Reinforcement Learning (RL):** Although not yet widely applied in biosensor diagnostics, RL holds potential for dynamic decision-making applications. An RL agent could, for example, learn to optimize which biosensor channels to prioritize in a multiplex system to achieve accurate diagnostics more efficiently. Flynn & Chang, (2024) suggest RL could also support adaptive assay control or real-time personalized intervention strategies.

A critical area of AI development for multiplexed biosensors is anomaly detection, particularly in health monitoring applications where the goal is to identify when a physiological parameter deviates from an individual's norm. Unsupervised and semi-supervised approaches such as autoencoders and one-class SVMs are well-suited to this task. These models learn a representation of “normal” based on an individual’s historical data and flag deviations as potential anomalies. For example, an autoencoder trained on typical glucose patterns from a specific user will reconstruct those inputs with low error. If the glucose curve becomes irregular, the reconstruction error increases, signalling a potential issue. This individualized approach reduces false positives by accounting for personal variability and avoids reliance on population-wide thresholds.

Another vital area is XAI, especially given the clinical settings in which biosensor data are used. Healthcare decision-making requires transparency and justification. Therefore, AI models must not only perform well but also produce interpretable outputs. In some cases, this can be achieved through inherently interpretable models like decision trees or rule-based systems. In others, complex models such as neural networks may be augmented with post-hoc explanation techniques that highlight which sensor features most influenced a given prediction or alert. This transparency is essential for clinician trust, facilitating the integration of AI tools as decision-support aids rather than opaque systems with unverifiable outputs.

Model lifecycle management is also a key consideration. Biosensor applications often involve long-term, real-world deployment, and AI models must maintain performance over time. However, as populations shift or device usage evolves, models can suffer from algorithmic drift. One solution is to design models that periodically retrain using new validated data to maintain accuracy. However, continuous self-learning AI introduces the risk of unmonitored behaviour changes, which can compromise safety and reliability. This has led to the emergence of algorithm vigilance practices—systematic monitoring of AI performance post-deployment, akin to pharmacovigilance in drug safety. In current practice, many medical AI systems are locked after initial training and only updated through controlled processes when sufficient new data accumulate, ensuring that updates do not compromise previously validated behaviour.

In summary, AI plays a central role in transforming multiplex biosensor data into meaningful, actionable information. From detecting anomalies tailored to individual baselines to providing interpretable results and maintaining robust performance over time, AI enhances the clinical and operational value of biosensor systems. As models grow more sophisticated, they promise earlier detection of health deviations, adaptive self-correction, and personalized analytics. Yet, the success of these systems depends not only on technical accuracy but also on transparency, fairness, and continual validation within their intended real-world contexts.

3.7. Use Cases & Applications

Multiplexed biosensors, in combination with Big Data infrastructure and AI, are transforming how health, environmental, agricultural, and public systems collect and act on data. These applications span personal, clinical, and societal domains. Each use case demonstrates how data flowing from individuals to populations must be carefully governed, with attention to ethics, privacy, and legal compliance.

3.7.1. Disease Outbreak Tracking

A timely and impactful application is real-time surveillance of infectious disease outbreaks. Multiplexed biosensors capable of detecting multiple respiratory pathogens (e.g. influenza, RSV, SARS-CoV-2 and its variants) can be deployed at various point-of-care settings such as clinics, pharmacies, or in at-home diagnostic kits. With patient consent and appropriate anonymization, results can be transmitted to public health surveillance platforms. AI systems analyse incoming data to generate high-resolution, real-time maps of disease activity (Jiménez-Rodríguez et al., 2022). This method significantly improves responsiveness over traditional surveillance systems, which depend on lagging case reports from clinicians.

During the COVID-19 pandemic, similar approaches were tested using rapid antigen tests and wastewater biosensors detecting viral RNA. If future multiplex tests can distinguish between variants or co-infections, they would support more targeted responses—for example, rapidly identifying the spread of a new strain or allocating vaccines based on local pathogen profiles.

In low-resource or remote regions, portable multiplex diagnostics can differentiate illnesses with overlapping symptoms—such as Ebola, malaria, and Lassa fever—enabling quicker containment. Researchers have also shown that combining biosensor outputs with external variables like climate and mobility data can improve predictive models of epidemic dynamics (Jiménez-Rodríguez et al., 2022).

Such use cases are permissible under the GDPR, which explicitly allows processing of personal data for public health in the public interest—provided the processing remains proportionate, secure, and transparent (European Parliament & Council of the European Union, 2016). As emphasized by Boronow et al., (2020), maintaining public trust requires clarity on how data are used and ensuring that pandemic surveillance respects individual rights.

3.7.2. Precision Medicine

Multiplexed biosensors also play a critical role in precision medicine—tailoring therapy based on real-time physiological monitoring. In oncology, for example, silicon biosensors can continuously track tumour biomarkers or chemotherapy metabolites in the bloodstream. These real-time data, combined with patient genomics and drug regimens, can be analysed by AI to dynamically optimize dosing, balancing efficacy and toxicity in a theragnostic framework.

In diabetes care, future multiplex wearables (e.g., microneedle patches) may monitor glucose, insulin, and cortisol simultaneously. AI algorithms could then make nuanced insulin dosing recommendations based on how an individual's metabolism responds throughout the day—beyond the scope of today's carbohydrate-counting models. Similarly, in autoimmune diseases such as lupus or rheumatoid arthritis, biosensors could track cytokine profiles to predict flare-ups before clinical symptoms emerge, enabling timely intervention.

Some EU-funded research projects are currently exploring AI-based biomarker monitoring for chronic inflammatory diseases. These systems typically begin at the personal level—sensor data are analysed locally and may provide real-time feedback (e.g., alerts for rising inflammation markers). Data may then be securely transmitted to clinicians or research platforms with appropriate consent. Aggregated datasets can inform subgroup identification and stratification for clinical trials or treatment refinement.

Privacy in precision medicine must be rigorously protected. Given the combination of biosensor data, genomics, and medical records, strong consent frameworks, end-to-end encryption, and opt-in secondary data use are necessary to comply with GDPR and medical confidentiality principles. However, the benefit is substantial: patients are empowered to participate actively in their care, and interventions can be more precisely aligned with their physiological state.

3.7.3. Environmental Exposure and Public Health

In environmental health, multiplex biosensors can greatly enhance urban air quality monitoring. For instance, sensor packs mounted on public buses can collect pollutant data across urban routes, producing dynamic, high-resolution pollution maps. AI can analyse these data streams in combination with health statistics (e.g., emergency admissions for asthma) to identify associations between pollution patterns and health outcomes.

This allows public officials to implement targeted measures—for example, restricting traffic or issuing alerts when certain pollutant combinations are predicted to elevate health risks. These insights rely on integrating open environmental datasets with sensitive health data. Aggregated health statistics (e.g., by neighbourhood or time) are often used to maintain privacy. Under European Commission (2022) provisions for secondary health data use, such analytics are allowed if data are anonymized or handled through secure infrastructure and appropriate approvals.

Municipal health departments or research institutes typically serve as neutral data stewards, ensuring responsible use. As the EHDS matures, such cross-domain research is expected to be streamlined, enabling continent-wide studies linking environment and health.

3.7.4. Agriculture and Food Quality

In agriculture, multiplexed biosensors support precision farming and animal health monitoring. Soil sensor arrays measuring multiple nutrients (e.g., nitrogen, phosphorus, potassium) can guide fertilizer application with spatial specificity. AI analyses these inputs to minimize waste and environmental runoff. Livestock wearables can track vital signs and stress levels, aiding in early disease detection and improving both productivity and animal welfare.

Sensor data are typically routed through farm management platforms, generating actionable insights. While privacy risks are lower than in healthcare, there are concerns regarding proprietary and worker-related data. To address these, initiatives such as the European agricultural data space aim to establish fair frameworks for data sharing. Data cooperatives and neutral brokers can ensure farmers retain control over their data while benefiting from collective insights (Joint Action Towards the European Health Data Space TEHDAS, 2023).

Although agricultural data are less regulated, ethical principles—transparency, equity, and non-exploitation—still apply. For example, it would be unethical for agribusinesses to use farmers' sensor data to manipulate market pricing.

3.7.5. Data Flow and Oversight: From Personal to Population-Level Insights

Figure 5 describes conceptually a tiered architecture for data governance and utility. At the base, biosensor data are collected at the individual level (e.g., from wearables, diagnostics, or environmental sensors). With user consent, these data are sent to secure platforms—personal apps for immediate feedback, and national or international databases for research and policy use.

AI and Big Data systems analyse these inputs to deliver two levels of insight: personalized feedback for individuals or their clinicians (e.g., alerts or tailored recommendations), and population-level intelligence for public health, urban planning, or scientific discovery. At all stages, oversight by neutral entities—data protection agencies, ethics committees, or regulatory bodies—is essential. These organizations act as data guardians, ensuring access is limited to authorized actors and aligned with legal and ethical frameworks.

An example: a wearable biosensor user contributes multi-analyte data, which their app uses to recommend actions like stress reduction. The anonymized version of that data is then pooled with national biosensor records, contributing to a study that discovers a new early warning sign for a chronic disease. The resulting algorithm update is pushed back to devices, improving detection accuracy for all users. Throughout, encryption, pseudonymization, and regulatory oversight ensure data use remains lawful and proportionate.

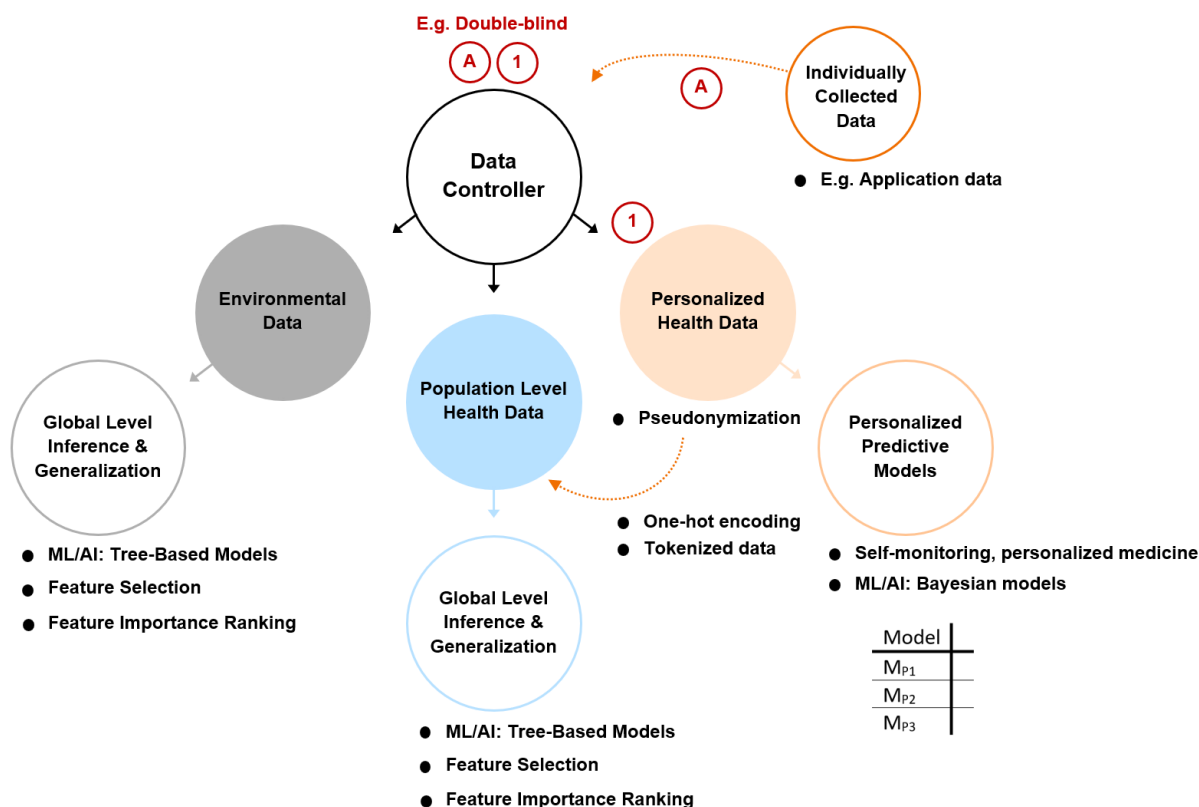


Figure 5. The Data Flow from Personal Biosensors to Population-Level Insights. This conceptual diagram illustrates how data from multiplexed biosensors—both health-related (e.g. wearable monitors, diagnostic kits) and environmental (e.g. air or water quality sensors)—move through secure integration platforms and are processed using AI and Big Data analytics. The resulting outputs serve two levels: (1) individualized feedback, such as alerts or treatment recommendations delivered via personal apps or healthcare providers, and (2) aggregated, anonymized insights for public health authorities, researchers, city planners, or policy makers. Oversight entities (e.g. ethics boards, health data authorities) ensure lawful access, compliance with regulations (like GDPR), and protection of individual rights at each stage. This layered data ecosystem supports both personal health empowerment and societal-level decision-making.

4. Conclusion

Multiplexed silicon biosensors integrated with Big Data and AI represent a frontier in both healthcare and environmental monitoring, promising faster diagnostics, personalized insights, and more proactive public health and environmental management. This report describes how these technologies function, their current applications and future potential, as well as the critical issues that accompany their deployment. Starting with an overview of multiplex biosensors in the introduction, we established how these technologies translate laboratory-grade diagnostic capabilities into portable, user-friendly formats. In the literature survey, we examined the current technological landscape and the diverse data streams generated—from digital POC devices to wearable environmental sensors—and discussed how these data sources contribute to Big Data ecosystems suitable for advanced AI-driven analysis. We also assessed the regulatory and ethical frameworks, with a particular focus on the European context, that shape the responsible collection, use, and sharing of biosensor-generated health and environmental data.

A throughout recurring theme is the opportunities and restraints going hand in hand. The capability to continuously measure and analyse multiple parameters can revolutionize POC diagnostics and epidemiology—enabling detection of disease outbreaks, tailoring treatments to individuals in real-time, reducing waste in agriculture, and responding dynamically to environmental threats. Yet, each of these advances requires careful design to avoid pitfalls: preserving privacy, ensuring data quality and interoperability, preventing algorithmic biases, and safeguarding against misuse of data or technology. The European perspective, reflected in instruments like GDPR, the AI Act, and the EHDS regulation, offers a blueprint for addressing many of these challenges by embedding ethical foresight and legal safeguards into the innovation process (European Commission, 2022; European Parliament & Council of the European Union, 2016). Concepts such as explicit consent, data minimization, fairness, explainability, and accountability need to be treated as non-negotiable design requirements for any Big Data or AI system using biosensor data (Sullivan, 2023).

One of the strengths of the EU approach is its emphasis on trust frameworks and data sovereignty. By granting individuals clear rights over their data and establishing trusted intermediaries, the EU seeks to foster an environment where citizens feel confident contributing data for societal benefit. This is crucial for novel technological solutions that otherwise remain limited. Public or neutral institutions play a vital role in this context by overseeing data governance and ensuring transparency, thereby acting as a barrier against unchecked data extraction by private entities. Moreover, Europe's leadership may shape international norms, promoting convergence toward globally respected standards for ethical, secure, and equitable data use.

We have also contrasted new biosensor-based practices with traditional ones in terms of ethical impact. Traditional health diagnostics often kept data local (e.g., a laboratory result on paper in a patient's file), thus posing less risk of large-scale privacy breaches. In contrast, modern digital biosensors broadcast data over networks, raising new privacy stakes. Likewise, environmental monitoring has traditionally focused on aggregate, population-level data, whereas emerging wearable

environmental sensors can link exposure data to specific individuals, locations, and behaviours. This shift significantly elevates the ethical stakes, as biosensors can now reveal continuous, intimate information about a person's body and lifestyle. These developments underscore the urgent need for robust frameworks to prevent potential harms such as biomarker-based discrimination or unauthorized surveillance. Ethical data stewardship must be a central pillar of technological deployment. The report underscores that the human element must remain central: the goal is to enhance human decision-making and welfare, not to replace human agency or exacerbate inequalities (Boronow et al., 2020; Sullivan, 2023).

Key Recommendations

- **Invest in interoperability and standards:** Global stakeholders should develop common technical standards for biosensor data formats and communication protocols. Regulatory harmonization across regions is also essential as biosensor data increasingly crosses borders. International consortia should align on device approval, consent procedures, and responsible data sharing practices.
- **Promote inclusive access:** To avoid deepening global health inequities, governments and international organizations should support access to multiplex biosensor technologies in low-resource settings. Subsidies, investment programs, and digital infrastructure are needed to ensure equitable deployment.
- **Continuous ethical oversight:** AI-integrated biosensor systems require ongoing monitoring. Ethics boards should routinely evaluate deployed systems for bias, unintended consequences, and alignment with ethical norms. Oversight must evolve alongside the technology and its social context (Kalodanis et al., 2025; Sullivan, 2023).
- **Individual rights:** Develop user-friendly tools that let individuals manage their data and consent preferences. Systems should be designed to respect rights such as human review of algorithmic decisions. Empowering users increases trust and encourages responsible data sharing.
- **Research and education:** Support research and development in privacy-preserving analytics like federated learning and homomorphic encryption. Educate healthcare professionals and the public to promote informed use and build trust in these emerging technologies. (Han et al., (2025) indicates movement in this direction, highlighting the feasibility of training AI models on distributed health data without pooling raw data.

In summary, the outlook for integrating Big Data and AI with multiplexed silicon biosensors is highly promising—provided development proceeds within a human-centric and policy-aware framework. European initiatives exemplify how technological innovation can be aligned with the protection of fundamental rights and democratic values. These efforts not only set a benchmark for responsible innovation within the EU but also have the potential to influence global practices as other regions observe and respond to their outcomes. For professionals operating at the nexus of data science and POC technology, the way forward requires more than just technical excellence. It demands proactive

collaboration with policymakers, ethicists, and the broader public to ensure that emerging solutions earn and sustain societal trust. The success of these technologies will ultimately be judged not only by their impact on health and environmental indicators, but also by how equitably, transparently, and ethically they are implemented.

As these systems evolve in the coming decade, their societal acceptance will hinge on whether they are perceived as fair, beneficial, and respectful of individual autonomy and dignity (Boronow et al., 2020). With thoughtful governance and inclusive design, multiplexed biosensors powered by AI have the potential to become foundational infrastructure in a more intelligent, resilient, and health-conscious society. Their development offers a valuable case study in how emerging digital innovations can be woven into the fabric of daily life—ideally, as models of ethically grounded, publicly trusted technological advancement.

References

- Beisenova, A., Adi, W., Janna Bashar, S., Velmurugan, M., B. Germanson, K., A. Shelef, M., & Yesilkoy, F. (2023). Machine-learning-aided multiplexed nanoplasmonic biosensor for COVID-19 population immunity profiling. *Sensors & Diagnostics*, 2(5), 1186–1198. <https://doi.org/10.1039/D3SD00081H>
- Bioensio. (2022, November 8). *Bioensio to develop novel, cost-effective methods for wafer scale biofunctionalisation in BioAssembler project*. <https://bioensio.com/bioensio-to-develop-novel-cost-effective-methods-for-wafer-scale-biofunctionalisation-in-bioassembler-project/>
- Boronow, K. E., Perovich, L. J., Sweeney, L., Yoo, J. S., Rudel, R. A., Brown, P., & Brody, J. G. (2020). Privacy Risks of Sharing Data from Environmental Health Studies. *Environmental Health Perspectives*, 128(1), 017008. <https://doi.org/10.1289/EHP4817>
- Bouderhem, R. (2024). Ethical and Regulatory Challenges for AI Biosensors in Healthcare. *Proceedings*, 104(1), Article 1. <https://doi.org/10.3390/proceedings2024104037>
- Brück, O., Sanmark, E., Ponkilainen, V., Bützow, A., Reito, A., Kauppila, J. H., & Kuitunen, I. (2024). European health regulations reduce registry-based research. *Health Research Policy and Systems*, 22(1), 135. <https://doi.org/10.1186/s12961-024-01228-1>
- DigiFinland Oy. (2024). *Tekoäly hyvinvointialueilla: Sosiaali- ja terveydenhuollon käyttötapaukset ja kansallinen edistäminen. Esiselvitystyön loppuraportti*. https://digifinland.fi/wp-content/uploads/2024/03/DigiFinland_tekoaly_loppuraportti_210324.pdf
- Dincer, C., Bruch, R., Kling, A., Dittrich, P. S., & Urban, G. A. (2017). Multiplexed Point-of-Care Testing – xPOCT. *Trends in Biotechnology*, 35(8), 728–742. <https://doi.org/10.1016/j.tibtech.2017.03.013>
- European Commission. (2019). *Ethics guidelines for trustworthy AI* (Shaping Europe’s Digital Future). <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

- European Commission. (2022). Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. *EUR-Lex*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197>
- European Commission, Directorate-General for Health and Food Safety. (n.d.). *My rights over my health data*. Retrieved 24 June 2025, from https://health.ec.europa.eu/ehealth-digital-health-and-care/my-rights-over-my-health-data_en
- European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. 1–88.
- Flynn, C. D., & Chang, D. (2024). Artificial Intelligence in Point-of-Care Biosensing: Challenges and Opportunities. *Diagnostics*, 14(11), Article 11. <https://doi.org/10.3390/diagnostics14111100>
- Gavrilăș, S., Ursachi, C. Ștefan, Perța-Crișan, S., & Munteanu, F.-D. (2022). Recent Trends in Biosensors for Environmental Quality Monitoring. *Sensors (Basel, Switzerland)*, 22(4), 1513. <https://doi.org/10.3390/s22041513>
- Han, G.-R., Goncharov, A., Eryilmaz, M., Ye, S., Palanisamy, B., Ghosh, R., Lisi, F., Rogers, E., Guzman, D., Yigci, D., Tasoglu, S., Di Carlo, D., Goda, K., McKendry, R. A., & Ozcan, A. (2025). Machine learning in point-of-care testing: Innovations, challenges, and opportunities. *Nature Communications*, 16(1), 3165. <https://doi.org/10.1038/s41467-025-58527-6>
- Jiménez-Rodríguez, M. G., Silva-Lance, F., Parra-Arroyo, L., Medina-Salazar, D. A., Martínez-Ruiz, M., Melchor-Martínez, E. M., Martínez-Prado, M. A., Iqbal, H. M. N., Parra-Saldívar, R., Barceló, D., & Sosa-Hernández, J. E. (2022). Biosensors for the detection of disease outbreaks through wastewater-based epidemiology. *Trends in Analytical Chemistry*, 155, 116585. <https://doi.org/10.1016/j.trac.2022.116585>

- Joint Action Towards the European Health Data Space TEHDAS. (2023, April). *Finland Country Visit Factsheet*. <https://tehdas.eu/app/uploads/2023/04/finland-country-visit-factsheet-04-2023.pdf#:~:text=unique%20identifier%20for%20linking%20individual,most%20of%20the%20national%20registries>
- Kalodanis, K., Feretzakis, G., Anastasiou, A., Rizomiliotis, P., Anagnostopoulos, D., & Koumpouros, Y. (2025). A Privacy-Preserving and Attack-Aware AI Approach for High-Risk Healthcare Systems Under the EU AI Act. *Electronics*, 14(7), Article 7. <https://doi.org/10.3390/electronics14071385>
- Laskey, R. (2022, March 8). What Is a Photonic Biosensor and How Will It Change PoC Diagnostics? *PhotonDelta*. <https://www.photondelta.com/news/what-is-photonic-biosensor-how-will-it-change-poc-diagnostics-market/>
- Prissel, C. M., Grossardt, B. R., Klinger, G. S., St. Sauver, J. L., & Rocca, W. A. (2023). Integrating Environmental Data with Medical Data in a Records-Linkage System to Explore Groundwater Nitrogen Levels and Child Health Outcomes. *International Journal of Environmental Research and Public Health*, 20(6), Article 6. <https://doi.org/10.3390/ijerph20065116>
- Puumala, L. S., Grist, S. M., Morales, J. M., Bickford, J. R., Chrostowski, L., Shekhar, S., & Cheung, K. C. (2023). Biofunctionalization of Multiplexed Silicon Photonic Biosensors. *Biosensors*, 13(1), Article 1. <https://doi.org/10.3390/bios13010053>
- Salvador, J.-P., Marco, M.-P., Saviozzi, G., Laschi, C., Arreza, F., Palacio, F., & Lopez, M. (2022). Portable flow multiplexing device for continuous, in situ biodetection of environmental contaminants. *Sensing and Bio-Sensing Research*, 37, 100505. <https://doi.org/10.1016/j.sbsr.2022.100505>
- Smith, A. A., Li, R., & Tse, Z. T. H. (2023). Reshaping healthcare with wearable biosensors. *Scientific Reports*, 13(1), 4998. <https://doi.org/10.1038/s41598-022-26951-z>
- Sullivan, M. (2023, October 20). *Key principles for ethical AI development*. Transcend. <https://transcend.io/blog/ai-ethics>

Suomi.fi. (n.d.). *Good practices for Service Developers—Using AI responsibly. How can you use AI in an ethical way?* <https://kehittajille.suomi.fi/guides/responsible-ai>

Xi, H., Jiang, H., Juhas, M., & Zhang, Y. (2021). Multiplex Biosensing for Simultaneous Detection of Mutations in SARS-CoV-2. *ACS Omega*, 6(40), 25846–25859. <https://doi.org/10.1021/acsomega.1c04024>