Summary of the context and overall objectives of the project

Fourth and final reporting period (1st July 2022 – 31st December 2023)

The digital revolution, in particular big data and artificial intelligence (AI), offers new opportunities to transform healthcare. However, it also harbours risks to the security of sensitive clinical data stored in critical healthcare information and communication infrastructure. Data exchange over the internet is perceived as impossible, posing a roadblock that hampers big data-based medical innovations. FeatureCloud's transformative security-by-design concept minimised the cyber-crime potential and enabled secure cross-border collaborative data mining endeavours. FeatureCloud was implemented as a community-extendible software platform to substantially reduce cyber-risks to healthcare infrastructure by employing the worldwide first privacy-by-architecture approach with two key characteristics: **(1) no sensitive data is communicated**, and **(2) data is not stored in one central point of attack**. Federated machine learning (FML) for privacy-preserving data mining, integrated with privacy-enhancing technology, like secure multiparty computation (SMPC) or differential privacy (DP), applies next-generation AI technology to medical purposes. Our ground-breaking cloud-AI infrastructure only exchanges learned model parameters, which are anonymous by default. Collectively, our highly interdisciplinary consortium covered all aspects of the value chain, from IT to medicine: Assessment of cyber-risks, legal considerations, management, international policies, federated AI technology, extendible App Store design and user interfaces, evaluation and translation in the clinic, and dissemination to the science community and the general public to maximise patient trust and societal acceptance of the new technology.

Conclusion: FeatureCloud's goals were bold and necessary, yet achievable, and our results paved the way for a socially agreeable big data era of the Medicine 4.0 age. Buttressed by 71 high-quality publications, our main product, the FeatureCloud.ai platform, offers an intuitive and user-friendly solution for the development and application of federated learning (FL) methods in biomedicine. Development templates and 55 apps for FL are accessible directly via the platform and have been tested through collaborations with real-world clinics.

Figure 1. *FeatureCloud's innovative concept revolutionises cloud communication for biomedical research because of its novel "Pirvacy by Architecture and Design" approach. While state-of-the-art security measures use policy and protective technologies to keep attackers from hacking into centralised "single-point-of-attack" data clouds, FeatureCloud's elegant federated solution has privacy and security designed into its very core architecture. Any and all sensitive data (e.g. personal or primary medical data of a patient) remain behind the safe firewalls of local hospitals or research institutions. Only abstract model parameters (that do not allow back-tracing to or identification of an individual patient) are communicated to FeatureCloud.*

Work performed from the beginning of the project to the end of the period covered by the report and main results achieved so far

We implemented and disseminated a stable production version of the FeatureCloud App Store including software development packages for corresponding computer-computer interfaces running as web servers - fostering an advanced, user-extendible App Store functionality. All source code is stored in Git repositories. Several hackathons, including one for external participants, supported app development and platform stress-testing. We have added SMPC and DP functionalities to secure the platform further. More than 50 apps are available now, from standard statistics apps to advanced apps for federated principal component analysis (PCA, see Hartebrodt et al. 2021 and 2022), artificial neural networks, random forest classifiers, and survival time prediction. In total, we prepared eight live demos to illustrate FeatureCloud's capabilities. Using the first FeatureCloud apps, we demonstrated the power of FML coupled with relevant privacy-enhancing technologies. We worked on typical medical application scenarios, beginning with the federated genome-wide association study (GWAS) tool 'sPLINK' (Nasirigerdeh R et al. 2022, published in *Genome Biology*) , which mimics the non-federated standard GWAS tool 'PLINK'. We demonstrated that (while currently available GWAS software, based on meta-analyses, loses accuracy when data suffers heterogeneously distributed outcomes or confounders) sPLINK gives the same results as the gold-standard tool 'PLINK'. 'sPLINK' implements federated Chi-squared tests as well as federated multimodal linear and logistic regression models. Likewise, we developed the first software for federated survival analysis, namely 'PARTEA' (Späth et al. 2022, published in *PLoS Digital Health*). It combines federated statistical modelling and differential privacy approaches based on Laplacian noise to generate privacy-preserving Kaplan-Meier plots. In addition, we developed, evaluated, and published 'flimma', a federated gene expression data analysis tool (Zolotareva et al. 2021, also published in *Genome Biology*). The FeatureCloud App Store itself was also evaluated and tested in relevant real-world scenarios (Matschinske et al. 2023, published in the *Journal of Medical Internet Research*).

Exploitation and real-world benefits of the project: The FeatureCloud platform and its underlying technology are versatile and can be applied across various markets and use cases. The demand for secure data accessibility and collaboration with external entities, while maintaining confidentiality, is substantial. Consequently, FeatureCloud.ai has been made available as an openly accessible platform. This will serve as a catalyst for app developers, start-up companies, and future projects.

Progress beyond the state of the art and expected potential impact (including the socio-economic impact and the wider societal implications of the action so far)

1) FeatureCloud contributed significantly to all three expected impacts mentioned in the work programme:

- **Improved security of Health and Care services, data and infrastructures**

  By addressing the evident roadblock in medical data mining – the need for centralised data mining for the development of high-quality AI tools but distributed clinical data – we improved the cyber security of computational health care services, patient data, and communication infrastructure. FeatureCloud's federated machine learning and privacy-enhancing technology engines erased the necessity to share sensitive data via a cloud.

- **Less risk of data privacy breaches caused by cyberattacks**

  FeatureCloud significantly reduced the risk of data privacy breaches by cyberattacks on health cloud services or communication channels between hospital and cloud. Instead of bringing the data to the AI, we bring the AI to the data.

- **Increased patient trust and safety**

  FeatureCloud technology ensures that no sensitive data is exchanged during the training of AI models. This strategy increases patient trust and safety significantly. The FeatureCloud platform complies with the EU's General Data Protection Regulation (GDPR) and Network and Information Security Directive (NISD) policies, respecting the criteria for software-supported medical devices of the European Medicines Agency (EMA) and the US-American Food and Drug Administration (FDA).

2) FeatureCloud furthermore contributed to the following most significant impacts not mentioned in the work programme:

- **The novel FeatureCloud technology will create new market opportunities.**

  FeatureCloud's App Store for client-side machine learning tools already had an enormous impact worldwide and will foster pan-European business development because of a huge emerging market in privacy-aware machine learning.

- **European society will benefit from new levels of personalised medicine, new possibilities for research of complex diseases like cancer, and lower costs of medical research.**

  FeatureCloud enables open science without boundaries, across domains and pan-European, generating new opportunities for research and the development of tools for personalised medicine as FeatureCloud apps address current privacy-, ethics-, and security-restrictions at the core. FeatureCloud thus lays the foundation to improve the overall quality of medical care while slowing down increasing healthcare costs in Europe and beyond.

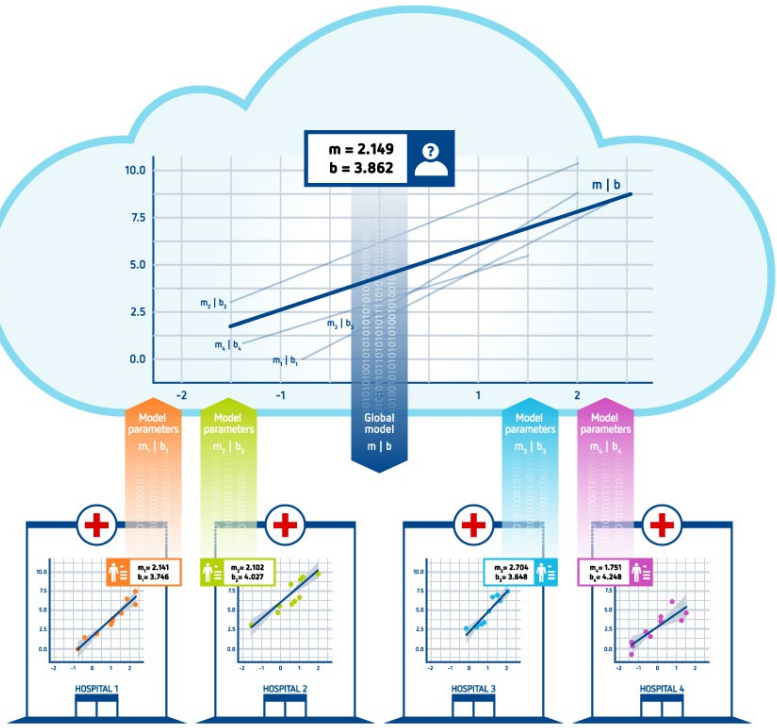

Figure 2. *Visualisation of what type of patient-based medical information is processed locally at each hospital (Hospital 1, 2, 3, and 4) and what type of abstract, privacy-preserving information is communicated to FeatureCloud. Only abstract, anonymous, and privacy-preserving model parameters are communicated (digitally transferred) to FeatureCloud, where the received information is, for example, used to form a higher-level, more accurate aggregated model.*

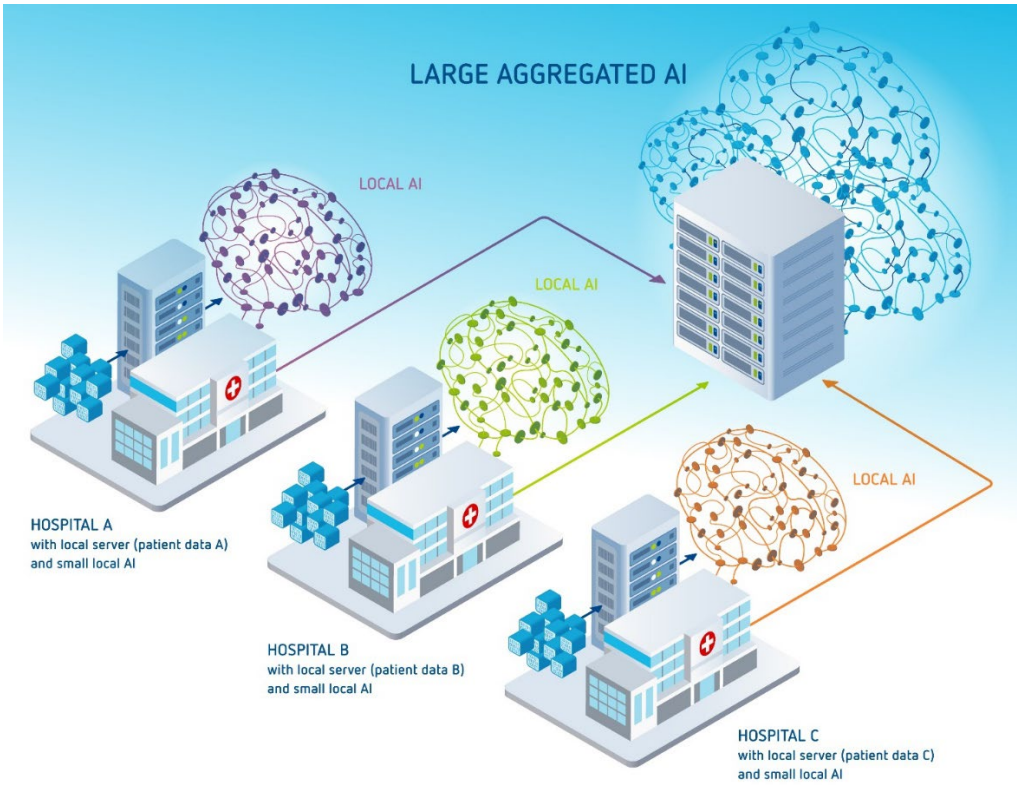

Figure 3. *Graphical representation of individual hospitals' local artificial intelligence (AI) and a higher-level, large aggregated AI that emerges as the result of a secure and privacy-preserving international research collaboration via the FeatureCloud platform and App Store. The figure shows what is communicated to the FeatureCloud platform and what happens with the communicated information. In detail: The purple, green, and orange "knowledge clouds" signify local bits of artificial intelligence (local AIs) that operate locally behind the safe firewalls of individual hospitals. The small, local AIs are securely generating AI models on local patient data, but only privacy-preserving model parameters (but no personal or primary medical data) are digitally communicated to FeatureCloud, where a large aggregated AI (blue) learns and constantly improves itself. The large aggregated AI constantly trains itself on novel input parameters – forming and continuously self-correcting (and therefore improving) a higher-level, aggregated AI model.*

Address (URL) of the FeatureCloud Platform and App Store

https://featurecloud.ai/

Address (URL) of the action's public project website*

https://featurecloud.eu/

**Please note that within the year 2024, the actions public project website will likely be moved to the above-listed featurecloud.ai domain. All content from the ".eu" project website will remain accessible on either website for at least 12 months after the end of the project.*

FeatureCloud Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078. This periodic summary report reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.