The digital revolution, in particular big data and artificial intelligence (AI), offer new opportunities to transform healthcare. However, it also harbors risks to the safety of sensitive clinical data stored in critical healthcare ICT infrastructure. In particular data exchange over the internet is perceived insurmountable posing a roadblock hampering big data based medical innovations. FeatureCloud's transformative security-by-design concept will minimize the cyber-crime potential and enable the first secure cross-border collaborative data mining endeavors. FeatureCloud will be implemented into a software toolkit for substantially reducing cyber risks to healthcare infrastructure by employing the world-wide first privacy-by-architecture approach, which has two key characteristics: (1) no sensitive data is communicated through any communication channels, and (2) data is not stored in one central point of attack. Federated machine learning (for privacy-preserving data mining) integrated with blockchain technology (for immutability and management of patient rights) will safely apply next-generation AI technology for medical purposes. Importantly, patients will be given effective means of revoking previously given consent at any time. Our ground-breaking new cloud-AI infrastructure only exchanges learned model representations which are anonymous by default. Collectively, our highly interdisciplinary consortium, from IT to medicine covers all aspects of the value chain: assessment of cyber risks, legal considerations and international policies, development of federated AI technology coupled to blockchaining, app store and user interface design, implementation as certifiable prognostic medical devices, evaluation and translation into clinical practice, commercial exploitation, as well as dissemination and patient trust maximization. FeatureCloud's goals are bold, necessary, achievable, and paving the way for a socially agreeable big data era of the Medicine 4.0 age.

We implemented a first FeatureCloud platform including corresponding computer-computer interfaces running as web server and fostering a basic app store functionality. It was developed using documented bi-weekly platform developer online conferences with all source code stored in Git repositories. Software development was organized into sprints and extended by two app development hackathons to test the overall system. Apps for basic federated computations of mean and standard deviation have been implemented first, as well as one enhanced app on federated principle component analysis (PCA). An app development and testing environment was deployed as well as a full documentation to aid app developers. In total, we also prepared five live demos to illustrate federated machine learning technology developed in the first period of FeatureCloud.

In addition to the first core apps of the FeatureCloud platform itself, we have worked on demonstrating the principle of federated machine learning. While the FeatureCloud prototype platform emerged, we have worked on stand-alone solutions for typical medical application scenarios. We began with a federated genome-wide association study (GWAS) tool: sPLINK, which mimics the non-federated standard GWAS tool PLINK (https://www.biorxiv.org/content/10.1101/2020.06.05.136382v1). We demonstrate that currently available distributed GWAS software (so-called meta-analysis tools) massively loses accuracy when the data suffers heterogeneously distributed outcomes or confounders. In contrast, sPLINK gives the exact same results as PLINK, and thus has the potential to become the new standard tool for genotyping in the future as it does not require any exchange of raw data between the participating institutions/hospitals and on top is not suffering any accuracy lost compared to the state-of-the-art centralized tools. sPLINK implements federated Chi-squared tests, as well as federated multimodal linear and logistic regression models. Likewise, we have developed first prototype software

for federated survival analysis: FedSurv. It combines federated statistical modelling and differential privacy approaches based on Laplacian noise to generate privacy-preserving Kaplan-Meier plots.

FeatureCloud contributes 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 – centralized data mining but distributed clinical data – we improve the cyber security of computational health care services, patient data and communication infrastructure by design and by architecture. FeatureCloud's federated machine learning engine erases the necessity to share sensitive data with a cloud.

• **Less risk of data privacy breaches caused by cyberattacks.**
FeatureCloud significantly reduces the risk of data privacy breaches caused by cyberattacks on health cloud services or on the 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.**
Based on trusted authority technology, like blockchains, we work on ensuring full control over the access rights to own sensitive data combined with the guarantee that no sensitive data is exchanged to learn the federated AI which could be traced back to individual patients (by design) will increase patient trust and safety significantly. Our FeatureCloud platform is in full accordance with E.U. GDPR and NISD policies, and it is developed with respect to the criteria for software-supported medical devices of the FDA and EMA, respectively.

FeatureCloud furthermore contributes to the following most significant impacts not mentioned in the work programme:

• **The novel FeatureCloud technology will create new market opportunities.**
FeatureCloud's replicability and scalability of client-side machine learning concepts will have an enormous impact worldwide and foster pan-European business, e.g. with spin-offs and start-ups because of a huge emerging market in privacy-aware machine learning.

• **The European society will benefit from new levels of personalized medicine, new possibilities for research of complex diseases like e.g., cancer, and lower costs of medical research.**
FeatureCloud enables open science without boundaries, cross-domain and pan-European, which will particularly allownew levels of cancer research, because FeatureCloud solves current privacy, ethical, security, and safety restrictions and will thus enable what was not possible to date, which can help to reduce increasing health costs in Europe by rising medical quality at the same time.