Scenario: Using data not created for research purposes in the context of collaboration with private companies

Sam is a researcher at a European public university examining the impact of diabetes on people's physical mobility. She designs a study investigating the influence of glucose levels in patients' blood on the type and intensity of physical activity in which they engage. To obtain relevant data, Sam reaches out to LargeTech, an American company responsible for the development of HealthyApp, a popular self-tracking app which allows users to manually log their health and activity data and to connect devices designed for automatic data collection. HealthyApp analyzes the data with the help of algorithms and provides users with insights as well as health and lifestyle recommendations.

The company sends Sam an anonymized data set sourced from users of continuous glucose monitors (CGMs), which are wearable devices allowing patients to obtain accurate, medical-grade and live readings of glucose levels in their bloodstream. In addition to data extracted from CGMs, the dataset also contains other information the app has collected on the users. This includes other health-related data, activity data from fitness trackers worn by some users in addition to CGMs, as well as various self-reported data points related to factors such as users' mood, weight, fertility, sleep patterns, calorie intake, etc. Although users are not explicitly informed that their data will be used in Sam's study, the terms and conditions of HealthyApp inform them that "their data might be shared and exchanged with LargeTech's partners and other third-parties for purposes such as research, development of new products and services, and refinement of existing products and services."

In exchange, LargeTech requests access to Sam's detailed research results and the right to review the manuscript before submission in order to ensure that it does not negatively impact the company's reputation, does not expose company's trade secrets and protects the anonymity of its users. Moreover, the company notes that it is unusual for them to share their data with independent researchers and ask Sam not to distribute the data set further. The study is a success, and a paper is published in a reputable journal following LargeTech's permission.

However, while investigating the dataset, Sam discovers an opportunity for authoring another paper. She notices a distinct correlation between users' blood glucose levels and their self-reported mood, and decides to investigate it further. She asks LargeTech for permission to use the dataset in the follow-up study but is denied as the company worries that some negative emotions experienced by HealthyApp's users could be associated with their usage of the app. Instead, Tom, the company's in-house researcher, suggests he and Sam join forces to conduct interviews with selected users about their experiences with CGMs and the devices' and data's impact on their mood. He also hints that Sam's participation in this study might open the possibility of her joining similar projects in the future.

Questions for researchers:

1. As a consequence of LargeTech's request not to distribute the dataset, is Sam able to guarantee that her research is transparent enough and potentially replicable by other researchers?
2. Does LargeTech's influence over the nature and potential publication of the follow-up study, allow Sam to conduct her work in a transparent, fair, full and unbiased way as required by ECCRI's honesty principle? Does the company's reaction to the proposed follow-up study have any bearing on the first study?
3. Considering that much of the data used by Sam has been self-reported by users of HealthyApp or sourced by non-medical-grade devices (e.g., fitness trackers), and has not been independently verified, is Sam able to ensure that the quality of the data set is high enough for it to be used in her study? Are consumer-grade data-collection devices, such as fitness trackers reliable enough for research purposes? Are medical-grade data-collection devices such as CGMs reliable enough for research purposes? Does the intermediation of the HealthyApp influence the reliability of the data?
4. Is it possible to identify conflicts of interests? How should Sam navigate these conflicts of interests (if any)?
5. Would Sam be justified in investigating the impact of HealthyApp on its users and publishing a paper on this subject despite LargeTech's lack of permission for such a study?
6. Should the possibility of joining LargeTech's future research projects influence Sam's decisions regarding the shape of the follow-up study? If yes, in which direction?

Questions for research administrators:

1. In collaborations between private companies and public institutions, what steps can be taken to ensure that profit motives do not negatively impact the overall quality of the study? Is the international nature of Sam's and LargeTech's collaboration relevant in this context?
2. Are the potential benefits arising from the collaboration (e.g., Sam's access to the dataset, LargeTech's greater insight into their app's impact on their users) distributed in a fair manner? Do all parties bear responsibility for risks associated with the study in a way that is proportional to their contribution and their perceived benefits?
3. What steps should the research administrators working with Sam take to ensure the integrity of the studies and to safeguard against LargeTech's undue influence over the results of Sam's work?
4. What should the research administrators working with Sam recommend to help the researcher in avoiding the risk of HARKing (Hypothesizing After the Results Are Known) on the basis of data obtained from LargeTech?

Questions for research ethics committees and research integrity offices:

1. Is the clause included in HealthyApp's terms and conditions enough to ensure the informed consent of data subjects involved in Sam's study? Does the transfer of data between the United States and Europe influence this determination in any way?
2. Are the research subjects involved in Sam's study adequately protected from potential harms such as deanonymization? Is there a difference between potential harms to research subjects likely to arise in the first study on diabetes and mobility, and those likely to arise in the second study investigating their emotional response to living with diabetes?

3. Would Sam's follow-up study adhere to the research design guidelines and procedures established at your institution? How should Sam approach this project to comply with best practices in research?
4. According to the guidelines established at your institution, does Sam's initial study involve human research subjects? Would it require clearance by your institutional ethics review board?