Briefing sheet and response sheet

Please study the scenario below.

Please put your responses in the answer boxes below the relevant question and when you have finished, submit this document to the assignment.

It is important to remember that this is an individual piece of work:

- It will be submitted to Turnitin to support the academic integrity of the assignment
- You must work on your own and must not involve anyone else in this assignment work.

Scenario

As artificial intelligence has been increasingly integrated into a whole new array of products in our home technology (e.g. AI assistants, doorbells, security camera systems etc.), smart homes are on the rise. Much of this technology is based on machine learning, a technique that uses large troves of data—including our voices, faces, homes, and other personal information—to train algorithms to recognize patterns. The most useful data sets are the most realistic, making data sourced from real environments, like homes, especially valuable. But to make these data sets useful for machine learning, individual humans must first view, categorize, label, and otherwise add context to each bit of data. This data annotation helps developers improve their devices, e.g. in terms of their navigation, object recognition, or speech recognition abilities.

The company selling these devices says that users have the opportunity to opt in or opt out to contributing such training data. However, consumers are not aware that human beings are going to be reviewing their raw audio/video footages.

The company also says that it shares only a subset of training footages with its data annotation partners and flags any images with sensitive information. The company further specifies, that when an image/audio footage is discovered where a user is in a compromising situation, including nudity, sexual interaction, or abusive interaction, it is deleted. It is not clarified whether this flagging would be done automatically by algorithm or manually by a person. In addition, the company's policy does not deem human faces as sensitive data, even if the people are minors.

Questions and answer boxes

(a)	Identify **ONE** ethical issue that is relevant to the scenario above and analyse it from a **Deontologist** perspective.

40 marks **Word limit: 350 words**	**Answer:** **Ethical Issue: Handling of Compromising Situations in User Data** The handling of compromising user data is a significant ethical issue because it touches directly on the principles of **privacy, dignity, and respect for individuals**. The company's policy states that footage containing compromising content (e.g., nudity, sexual, or abusive behaviour) is flagged and deleted, but it is unclear whether this is done automatically or manually. This ambiguity suggests potential privacy breaches, making it a critical focus from a deontological perspective.. **Ethical Considerations:** From a deontological standpoint, the primary focus is on the duties that the company has towards its users and whether these duties are upheld irrespective of the outcomes. The key duties in this context include: • **Duty to Protect Privacy**: The handling of raw footage, without a clear commitment to automated flagging and deletion, poses an ethical risk. The company's responsibility is to ensure that no human ever views sensitive content unless absolutely necessary. Allowing human annotators to access this type of footage, even with a flagging system, risks breaches of privacy that are intrinsically unethical, as it treats individuals' personal moments as subjects for external scrutiny. • **Duty to Treat Users with Dignity and Respect**: In deontological ethics, it is essential to treat individuals as ends in themselves, not as mere tools for achieving goals. When a company allows human access to sensitive data without the users' explicit understanding and consent, it compromises the dignity of these individuals. This is particularly troubling when such footage involves compromising situations, where the individuals are likely unaware that their private moments could be subject to human evaluation. • **Duty of Transparency**: The policy's ambiguity regarding whether sensitive footage is flagged automatically or by a human reviewer compromises the duty of transparency. Users must be fully informed that their data, especially compromising footage, might be viewed by human beings. The failure to clarify this process amounts to an ethical breach of honesty and openness. Deontological ethics holds that actions must align with moral principles, such as truthfulness, regardless of outcomes. The company's lack of clear communication breaches its duty to be forthright.
(b) **30 marks**	Using the **Ethical OS toolkit**, identify **ONE** risk zone that is most applicable to the case study above with its related unintended consequences. Justify the reasons for your choice.

Word limit: 250 words	**Answer:** One of the risk zone I believe is the most applicable to the case study above is the Risk Zone 6: Data Control and Monetization. **Justification:** **Risk Zone 6: Data Control and Monetization** applies to the scenario involving AI-based home technology. The Ethical OS toolkit emphasizes the importance of transparency and user control over personal data. The company's unclear practices around data collection and human review raise concerns about how user data is managed and shared without explicit consent, aligning with this risk zone. **Structured Analysis Based on the scenario:** **Transparency in Data Collection**: Users can opt in or out of data sharing but are not clearly informed about human review. The checklist question, "How can you be more transparent about this?" applies here. Explicitly disclosing data review practices is crucial to ensure informed consent. **User Control and Rights**: Users should have the right to access and control data collected about them. The scenario's lack of clarity around sensitive data handling highlights the need for users to know and manage their data, ensuring autonomy. **Profit from User Data**: While the company benefits from sharing data with partners, users are neither informed nor share in potential profits. This imbalance raises ethical concerns about user consent and fair benefit. **Associated Risks:** • **Privacy Breach**: Sensitive data mishandling or leaks jeopardize user privacy. • **Erosion of Trust**: Lack of transparency can damage user trust. • **Monetization without Consent**: Users should consent to and benefit from any profit made from their data.
(c)	Identify the subsection of Principle 1 (**Public Interest**) of the **BCS Code of Conduct** that is the most relevant to this scenario. Justify the reasons for your choice.

30 marks **Word limit: 250 words**	**Answer:** The most relevant subsection of **Principle 1 (Public Interest)** of the BCS Code of Conduct for this scenario is **1(a)**: *"have due regard for public health, privacy, security and wellbeing of others and the environment."* **Justification:** Subsection 1(a) directly relates to the handling of sensitive user data by the company. The scenario involves AI home technology that collects personal data, including audio and video footage, which may be subject to human review. This raises significant concerns about user **privacy, security** and **well-being**, which are core aspects of subsection 1(a). The ambiguous policy on whether footage is reviewed by humans potentially jeopardizes users expectation of privacy and can lead to security risks if sensitive data is mishandled or leaked. **Structured Justification:** • **Privacy Concerns**: Subsection 1(a) emphasizes the importance of protecting public privacy. The scenario demonstrates that the company's unclear disclosure about human data review undermines user privacy and violates their right to be informed about who has access to their personal information. • **Well-being and Security**: The well-being of users can be compromised when sensitive data is at risk of being exposed to human annotators or potentially leaked. Ensuring user data is handled securely aligns with the requirement to protect public security and well-being. • **Public Trust**: Upholding public interest by safeguarding privacy and ensuring transparent data practices enhances public trust. If not adhered to, breaches of privacy could result in loss of trust and reputational damage, impacting the broader public perception of technology and IT practices.