Data Privacy Impact Assessment

Submitting controller details

Name of controller	Heidi Health
Subject/title of DPO	Heidi Scribe
Name of controller DPO	Yassin Omar Head of Compliance yassin@heidihealth.com

Why is a DPA Needed?

Heidi processes and transcribes clinical conversations, capturing details like different speakers, medical terminology, and symptomatology. From this, a clinical note is generated. The clinician can also generate clinical documents, such as referral letters and patient explainer documents. These documents will follow templates already defined by Heidi, or the clinician can create their own template.

This system is designed to alleviate the administrative burden on healthcare professionals, allowing them to focus more on patient care rather than paperwork. The Heidi Scribe will leverage natural language processing (NLP), speech recognition technology, and machine learning algorithms to understand and interpret complex medical dialogue, identify key health information, and categorise data into the appropriate sections of an EHR.

The need for a DPIA was identified due to the sensitive nature of the data involved and the potential risks to patient privacy. Medical conversations contain highly confidential information, and the processing of this data through AI technologies can raise privacy and security concerns. The DPIA is necessary to evaluate these risks, ensure compliance with data protection laws and implement appropriate safeguards to protect patient information.

Describe the Nature of the Processing

Heidi is a healthcare IT system, specifically a cloud-based artificial intelligence medical scribe platform. It is a standalone software that is used to generate comprehensive clinical documentation using a combination of speech-to-text software, note taking and artificial intelligence models. Heidi is accessible via desktop and mobile browser to registered users, with servers and data hosted locally in the UK for all UK users.

Heidi works by transcribing speech into text from a healthcare encounter such as conversations between clinicians and patients or by clinicians dictating their clinical findings, impression and/or management plans before, during and after the healthcare encounter. The clinician can also add additional contextual notes about the healthcare encounter which they may not wish to verbalise during the healthcare encounter. The clinician is also able to set and modify various settings within the Heidi platform in order to customise their Heidi experience as well as how their clinical documentation is structured and written. To generate the requested clinical documentation, the transcribed text and contextual notes along with the various user controlled settings are then through an artificial intelligence model which then generates the requested clinical documentation based on the data that has been given to the AI model.

The comprehensive clinical documentation generated by Heidi can then be copied or integrated into an electronic medical record system or used with other word processing or communication tools to provide other clinicians and/or the patient with relevant information related to the healthcare encounter and the patient's care.

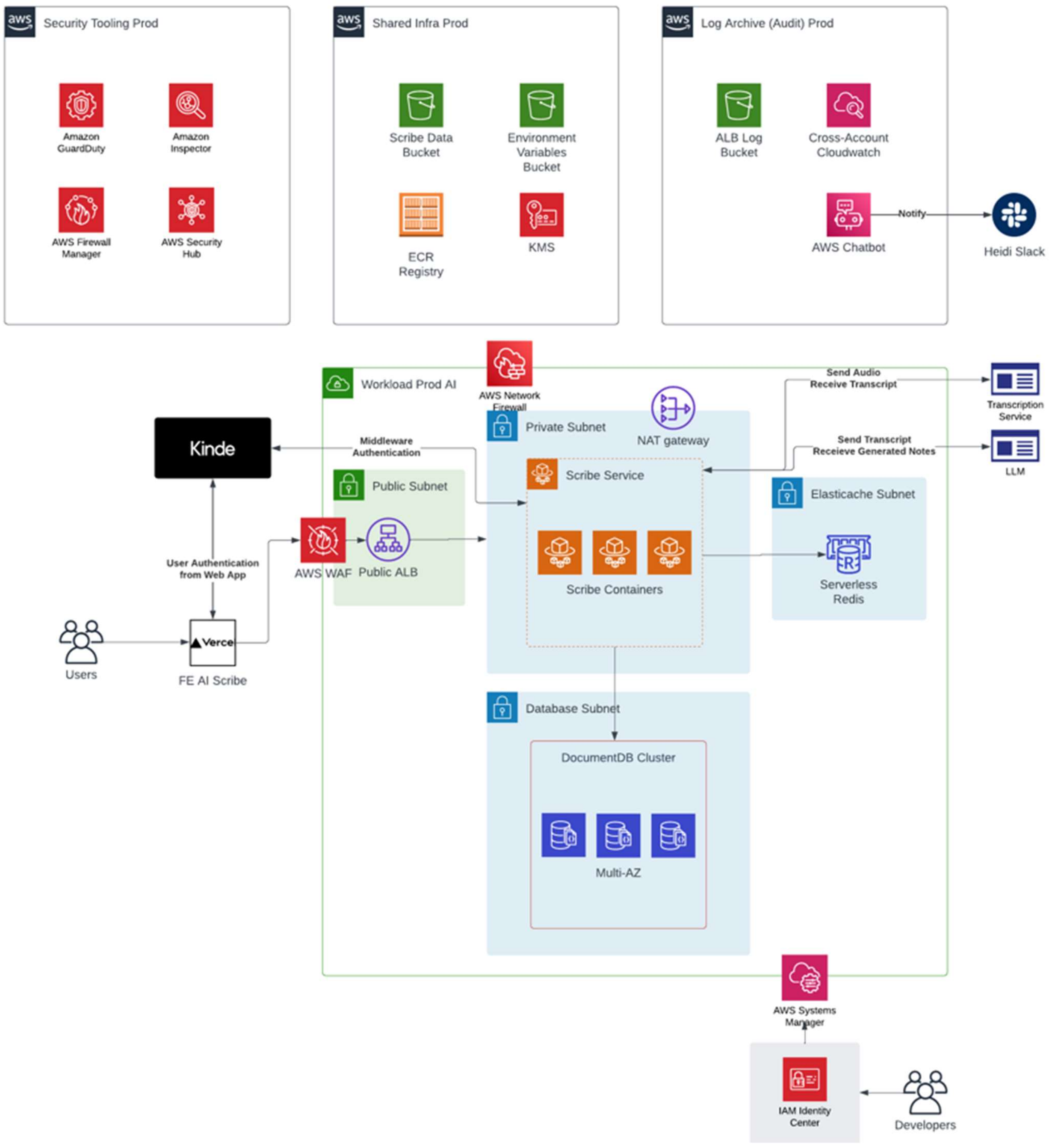
The intended use and recommendations for Heidi are as follows:

- Heidi should be used by qualified and registered clinicians to assist them in writing their clinical documentation.
- Heidi should not be used as a clinical decision making tool and is not a substitute for medical assessment.
- Heidi's generated clinical documentation is intended to reduce the amount of time it takes clinicians to complete their medical records; however the clinician is ultimately responsible for their clinical documentation, and must ensure that the content of the notes and documents accurately reflects the healthcare encounter for which the documentation has been generated.

Heidi is noted to have the following limitations that users must be aware of:

- Heidi's generated clinical documentation is based on the clarity and quality of the speech & text data that is provided in the healthcare encounter. Users must review all clinical documentation generated to confirm their accuracy before capturing it in their electronic medical records or distributing documentation to other clinicians and/or patients.
- Hardware issues such as poor microphone quality may cause sub-par audio being captured, resulting in an inaccurate text transcript which does not adequately reflect the healthcare encounter information. Users are recommended to test the quality of their microphones prior to and while using Heidi.
- An unstable or slow internet connection may result in delays in information processing and potentially not capturing some or all of the healthcare encounter information. Therefore it is vital that users ensure they have a stable and fast internet connection when using Heidi.
- Heidi's AI models occasionally make mistakes which may not accurately reflect the information discussed in the healthcare encounter. Users must ensure they have reviewed all clinical documentation generated by Heidi to confirm their accuracy before importing it in their electronic medical records or distributing documentation to other clinicians and/or patients.

Software Architecture & Data Flows



Describe the scope of the processing

The scope includes recording, transcribing, and generating summaries of clinical consultations. It encompasses the entire duration of the consultation, covering various medical details, additional clinician notes, and dynamic command line inputs that are inputted by the clinician. Special category data including information regarding patient health, will be de-identified and pseudonymized before the transcript is processed into clinical notes or other clinical documents. Personal information of the patient may also be collected by the clinician - this includes the patient's name, gender, and date of birth. Patients may be under the age of 18.
This processing will occur every time a clinician performs a session on Heidi - in other words, every time they click 'start recording'. This data will cover both data from the patient themselves (though it is de-identified) and the clinician too - such as their templates, note-taking style, clinician type, email address etc. No identifiable recordings are stored, or will be accessible. Clinicians retain ownership of all transcripts, clinical notes, and clinical documents and can decide how long this data is stored. Additionally, the patient information contained in these transcripts and clinical notes/documents will only be accessed externally for the purpose of troubleshooting with the express permission of clinicians.

Describe the context of the processing

Heidi operates within the context of medical consultations, capturing and processing information in the healthcare setting. It is designed to improve clinical documentation and streamline healthcare practices. Clinicians using Heidi are familiar with our processing requirements, as detailed in our Terms of Use, Privacy Policy, and Safety & Security page on our website. This ensures the information regarding the use of patient data is easily accessible to them. Clinicians have the right to access, update, and delete transcripts, clinical notes, and clinical documents at their discretion.

There have been concerns over the processing and security of AI scribes in clinical settings. Key issues include:

- **Data Privacy:** Protecting patient data is crucial due to the sensitive nature of medical information. Ensuring compliance with regulations of local jurisdictions is essential.
- **Data Security:** Ensuring the security of stored and transmitted data to prevent unauthorised access, breaches, and leaks.
- **Accuracy:** Ensuring that the AI accurately transcribes consultations and generates clinical notes without errors, which could potentially lead to misdiagnosis or incorrect treatment.
- **Bias:** Addressing potential biases in AI algorithms that could lead to disparities in treatment based on patient demographics.
- **Ethical Use:** Addressing ethical concerns regarding the use of AI in healthcare, particularly around patient consent and the potential for reduced human oversight in clinical decision-making.

To address these security and processing concerns, Heidi employs stringent privacy and security measures, including end-to-end and at rest encryption, regular audits, real-time security monitoring, and penetration testing. We hold ISO27001, Cyber Essentials and SOC2 certifications, align with GDPR and other international data privacy regulations. Accuracy, potential bias, and ethical use concerns are addressed in our Clinician Terms of Use and Usage Policy, which outline the clinician's responsibility to obtain consent in line with their local jurisdictions regulations, and review all outputs for accuracy and quality.

Additionally, Heidi determines the type and level of access granted to individual users based on the "principle of least privilege." This principle states that users are only granted the level of access absolutely required to perform their job functions. Permissions and access rights not expressly granted shall be, by default, prohibited. Heidi's primary method of assigning and maintaining consistent access controls and access rights shall be through the implementation of Role-Based Access Control (RBAC). Wherever feasible, rights and restrictions shall be allocated to groups. Individual user accounts may be granted additional permissions as needed with approval from the system owner or authorised party.

AI scribes for transcribing clinical consultations are relatively novel and represent a significant advancement in healthcare technology. Their novelty lies in their high-quality voice recognition technology that accurately captures medical jargon, accounting for multiple speakers and varying accents, their ability to transcribe consultations in real time, and Advanced NLP algorithms that can understand and process medical terminology and context accurately.

Describe the purposes of the processing

The primary purposes include improving clinical documentation, aiding healthcare professionals in note-taking, and generating consult summaries. Our technology enables clinicians to focus on patients during the consultation, contributing to improved patient care. It also acts as a valuable tool for medical practitioners, saving them hours of administrative time per week. We also use aggregated de-identified information from these consults to improve our models and outputs, ultimately improving both patient care and clinician experience.

To ensure the highest level of accuracy and reliability in our models, we implement rigorous processes for eliminating biases from our outputs. This involves continuous monitoring and evaluation of our algorithms to detect and mitigate any potential biases that may arise. Our approach includes diverse and representative data sets, regular audits, and feedback loops from clinicians to refine and enhance the system.

Furthermore, we employ robust governance principles to maintain the integrity and trustworthiness of our technology. This includes sophisticated encryption protocols, de-identification of date, real-time safety monitoring, regular system audits, and penetration tests. By upholding these governance standards, we ensure that our technology not only meets but exceeds industry expectations, providing a safe and effective tool for healthcare professionals.

Health Information Listing

The following patient data items may be recorded and transcribed by Heidi:

- Name
- Address
- Phone number
- Gender
- Sexual orientation
- Date of birth
- Relationship status
- Family and social history
- Medical history
- Progress notes
- Medications & prescriptions
- Allergies
- Diagnosis status
- Lab orders & results
- Disability

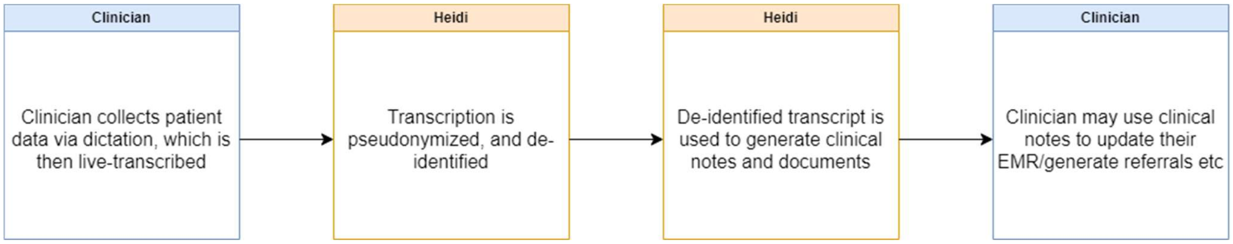
Personal Information Flow Table

Description	**Type**
Clinician obtains consent to record consult, which is then live-transcribed	Collection
Processing transcript through pseudonymization[1] and de-identification models	Use
De-identified transcript processed to generate clinical notes and documents	Use
Clinician may use clinical notes and documents to update their EMR/generate referrals etc	Use

1. Pseudonymization is the process of transforming personal data in such a way that individuals cannot be identified without additional information. This is done by first identifying sensitive data types, including transcripts, patient information, clinician notes, and generated notes. Sensitive data is encrypted both while in transit, and at rest, and all keys are managed securely. Machine learning (ML) techniques are used to de-identify transcript data, targeting entities like names, genders, addresses, emails, and phone numbers This is done by replacing identifying fields within a data record with artificial identifiers, or pseudonyms. For

example, instead of storing a person's name, the data might store a unique code that only authorized personnel can trace back to the original individual. The key to re-identify the data is kept separately and securely, ensuring that even if the pseudonymized data is accessed, the privacy of individuals remains protected. This technique is crucial for maintaining privacy while still allowing data to be used for analysis and improving ML accuracy.

Personal Information Flow Diagram


Clinician
Clinician collects patient data via dictation, which is then live-transcribed
Heidi
Transcription is pseudonymized, and de-identified
Heidi
De-identified transcript is used to generate clinical notes and documents
Clinician
Clinician may use clinical notes to update their EMR/generate referrals etc

Consider how to consult with relevant stakeholders

Consulting with relevant stakeholders is a critical step in ensuring the successful development and deployment of Heidi and we try to balance innovation with privacy and ethical considerations. Our approach to stakeholder consultation is multifaceted, designed to capture a broad spectrum of insights and concerns including:

Healthcare professionals:
Early and ongoing consultation will take place through surveys, interviews, and focus groups. Initial consultations will aim to understand their expectations, concerns, and suggestions regarding the use of AI in medical documentation. This will occur at the early development phase to guide the project's direction and once again during the pilot phase before full-scale implementation.

Cross functional teams:
Development of Heidi involves a cross-functional team, including product managers, AI developers, engineers, compliance officers, and user experience (UX) designers. Each plays a crucial role in ensuring the project meets its objectives while addressing privacy and ethical considerations.

Consulting experts:

Given the sensitivity of the data involved, consulting with information security experts is essential to assess and mitigate risks related to data breaches, unauthorised access, and data integrity. Their expertise will guide the implementation of robust security measures.Healthcare IT experts insights will ensure that the Heidi Scribe system integrates seamlessly with existing healthcare IT ecosystems, including EHR systems, and adheres to industry standards and best practices.

Describe compliance and proportionality measures

Lawful Basis for Processing:

Below is an overview of the personal data we collect and the legal basis for processing it. When processing data based on legitimate interests, we apply the following criteria:

- Purpose – is the purpose of processing personal data legitimate?
- Necessity – Is the processing necessary to achieve that purpose?
- Balance – do the individual's interests, rights or freedoms override the legitimate interest?

The collection of personal data via our website serves the following purposes:

- When providing information about products and services you requested through email subscriptions, the lawful basis for processing is consent.
- When providing related information about identified areas of interest, the lawful basis is legitimate interests.
- When enabling individuals to exercise their rights over personal data, the lawful basis is compliance with a legal obligation

Organisations such as Heidi which provide services to the NHS, may rely on the following legal bases for processing data:

- To support data processing for NHS services, using legal powers provided by:
  - The National Health Service and Community Care Act 1990
  - The NHS Act 2006
  - The Health and Social Care Act 2012
- To process personal and special category data in accordance with UK GDPR:
  - Art. 6(1)(e) – Public task
  - Art. 9(2)(h) – healthcare purposes
- To process personal data in line with the Data Protection Act 2018:
  - Condition 2 of Schedule 1 – Health and Social Care Purposes.
- To handle confidential information under the Common Law Duty of Confidentiality:
  - Implied consent - Heidi also does not undertake any automated decision-making or profiling in relation to your personal data as defined under Article 22 of the GDPR. Personal and special category data processed by Heidi will be retained in line with the retention periods specified in the "NHS England Records Management Code of Practice."“

Data Processing Purpose: The processing of data through Heidi is strictly aligned with the purpose of enhancing clinical documentation and facilitating efficient healthcare practices. By accurately transcribing clinical consults, Heidi helps to streamline documentation processes, allowing clinicians to focus more on patient care. Regular audits and assessments ensure that the data processing is achieving its intended purpose effectively and efficiently.

Alternative Solutions: While manual transcription and traditional documentation methods could achieve similar outcomes, they are often time-consuming and less efficient. Heidi offers a real-time, accurate, and efficient solution, providing a significant improvement in clinical workflows and reducing administrative burdens on healthcare providers.

Preventing Function Creep: To prevent function creep, Heidi's data processing activities are strictly governed by clearly defined use cases and regular reviews. Any expansion of data usage or new features undergoes a rigorous evaluation process to ensure alignment with the original purpose and compliance with regulatory requirements.

Data Quality and Minimisation: Data quality is ensured through continuous monitoring and validation processes. Only the minimum amount of data necessary for the specified purpose is collected and processed. Regular data audits and updates help maintain accuracy and relevance, ensuring compliance with the principle of data minimisation.

User Information and Rights Support: Users are provided with comprehensive information about data processing activities through an accessible web page detailing our Privacy Policy and Terms of Use. This includes information on their rights to access, correct, delete, or restrict their data. We also offer support channels to assist users in exercising their rights and addressing any concerns.

Ensuring Processor Compliance: We enter into strict data processing agreements with all third parties involved in the handling of data (e.g our cloud hosting provider for storing transcripts and clinical notes). These agreements are designed to ensure that no user data can be accessed, used, or stored by third parties beyond what is necessary for the specific purpose for which it was shared. We also enforce zero retention policies with all third-party service providers. This means that after the necessary data processing tasks are completed, no data is retained, ensuring that your clients' information cannot be reused or accessed for any other purpose.

In conclusion, Heidi takes comprehensive measures to ensure compliance with regulatory bodies, emphasising transparency, data security, and user rights. Our robust policies and practices are designed to protect patient information and maintain trust in our data processing activities.

Risk Identification

Source of Risk	**Likelihood of harm**	**Severity of harm**	**Overall risk**
1: Company systems are data are breached by unauthorised persons or becomes unavailable	Remote	Severe	Medium
2: Company data is breached due to human error and/or misunderstanding of company requirements	Remote	Severe	Medium
3: Staff are not aware of security threats and best practices resulting in a compromise of company systems and data	Remote	Severe	Medium
4: Company systems and data are breached in transit due to improper encryption	Remote	Severe	Medium

5: Company systems and data are breached in a non-production environment	Remote	Severe	Medium
6: System is unavailable during critical times	Remote	Severe	Medium
7: Speech-to-text model fails to accurately transcribe medical terminology, drug names etc	Somewhat Likely	Moderate	Medium
8: Public facing Privacy Policy (3) is inaccurate, omits critical information, and/or is out of date.	Somewhat Likely	Severe	High
9: Consent for processing of PII is not captured and can't be demonstrated when needed.	Remote	Severe	Medium

Risk Reduction

Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5				
Risk	**Options to reduce or eliminate risk**	**Effect on risk**	**Residual risk**	**Measure approved**
1	Strict physical access controls	Reduced	Low	Yes
2	Regular policy and security training	Reduced	Low	Yes
3	Regular policy and security training	Reduced	Low	Yes
4	Adequate encryption process in place	Reduced	Medium	Yes

5	Strict data access controls and admin roles	Reduced	Low	Yes
6	Redundant systems	Reduced	Low	Yes
7	Using speech to text model trained on medical data	Reduced	Low	Yes
8	Internal systems to ensure public facing Privacy Policy is regularly updated	Reduced	Medium	Yes
9	Process to obtain consent outlined in Clinician Terms of Use	Reduced	Low	Yes

Privacy Breaches

If an Heidi Health Trading Pty Ltd employee, contractor, user, or customer becomes aware of an information security event or incident, possible incident, imminent incident, unauthorised access, policy violation, security weakness, or suspicious activity, then they shall immediately report the information to the Compliance Lead or CTO. All relevant authorities, and stakeholders will be notified of the privacy breach as soon as possible.

For critical issues, the response team will follow an iterative response process designed to investigate, contain exploitation, eradicate the threat, recover system and services, remediate vulnerabilities, and document a post-mortem with the lessons of an incident.

Summary of Incident Response Process:

- Event reported
- Triage and analysis
- Investigation
- Containment & neutralisation (short term work)
- Recovery & vulnerability remediation
- Notify any relevant external stakeholders and authorities
- Hardening & Detection improvements (lessons learned, long term work)

Sign Off and Record Outcomes

Item	Name/position/date	Notes
Measures approved by:	Yassin Omar/ Data Protection Officer/ 12 April 2024	Integrate actions back into project plan, with date and responsibility for completion
Summary of DPO advice: Continue to manage risks and assess any new risks especially regarding data privacy and security protocols		
DPO advice accepted or overruled by:	Raghav Sharma	
Comments: DPO advice accepted		
This DPIA will kept under review by:	Yassin Omar	