TheHill & Oxford University Hospitals NHS Foundation Trust

Ambient Voice Technology (AVT) in Acute Services - An Evaluation

Foreword

The rapid evolution of artificial intelligence in healthcare presents both an opportunity and a responsibility for NHS Trusts. As clinical, operational, and digital pressures continue to grow across the NHS, we must ensure that emerging technologies are explored with rigour, transparency, and a clear alignment to clinical need. Our Ambient Voice Technology (AVT) pilot was undertaken with precisely this commitment in mind.

This work by TheHill team reflects one of the most comprehensive evaluations of AVT undertaken in an acute NHS setting to date. It places clinicians' lived experience at its centre, recognising that meaningful digital transformation is only achieved when technology genuinely improves the working lives of staff and supports safer, more effective patient care. What is striking in these findings is the consistency of the message from frontline colleagues: AVT has the potential to significantly reduce administrative burden, improve wellbeing, and enable more focused and compassionate patient interactions.

I am grateful to my clinical colleagues who contributed their time, insight, and honest reflections. Their openness has ensured that our next steps will be informed by real-world experience rather than assumptions. As we consider future procurement and deployment, our focus will remain on supporting staff wellbeing, clinical quality, and sustainable service delivery. This evaluation gives us confidence to move forward—cautiously, responsibly, and with optimism about the role that well-governed AI can play in easing the pressures on an overstretched system.

Dr. Ben Attwood, Chief Digital and Information Officer, Oxford University Hospitals NHS Foundation Trust

Innovation in the NHS is most powerful when it grows directly from the realities of frontline practice and is supported by leadership who recognise its strategic value. The decision to pilot Ambient Voice Technology (AVT) emerged from exactly this combination. Colleagues across the Trust had been voicing, with increasing urgency, the pressure of rising administrative burden—the late-evening letter writing, the documentation that spilled beyond clinic hours, and the mental effort required to keep pace with busy, complex workloads. Many clinicians, from a wide range of specialties, were keen to trial a tool that offered a genuine alternative. At the same time, national policy has highlighted the importance of exploring technologies that can lighten this load safely, and our Board recognised the need to test AVT carefully and responsibly within real clinical settings.

This report reflects the voices of more than a hundred OUH colleagues who tested AVT in highly varied clinical settings. Their feedback, recorded through logs, surveys, and interviews, has been honest, nuanced, and grounded in the complexity of real practice. The results are clear: when AVT is matched to the right workflow, it can be transformative. Staff describe reduced stress, improved work–life balance, and the ability to focus more fully on the human interaction at the heart of clinical care. These psychological and organisational benefits matter deeply—they influence retention, morale, and the quality of care delivered.

Equally important are the areas where AVT struggled. Complex, multi-voice, or highly qualitative environments challenged all solutions tested, and there remains essential work to do on template design, governance clarity, and integration. Innovation is rarely linear; it requires curiosity, critical reflection, and a willingness to iterate. This evaluation gives us the insights needed to take the next steps responsibly.

I would like to thank all the clinicians, nurses, allied health professionals, and operational colleagues who took part, as well as the teams within TheHill and our digital directorate who supported the project. Particular thanks to Ravi Chal, Ed Jaspers and Ian Pinnell, who were instrumental in designing, delivering and reporting on this pilot. Your commitment to improving services for patients—and to finding healthier ways of working—has shaped this report and provided us with much needed information for future decisions.

Megan Morys-Carter, Director of Digital Innovation, Oxford University Hospitals NHS Foundation Trust, Director of TheHill

Contents

1. Executive Summary

1.1 Project Overview

The Ambient Voice Technology (AVT) pilot at Oxford University Hospitals NHS Foundation Trust was launched to address the increasing administrative burden faced by clinical staff and to explore how AI-enabled voice tools could improve both productivity and patient care. The project was broad in scope, involving a diverse range of departments and clinical settings, from highly standardised outpatient clinics to more complex environments such as Critical Care and Community Paediatrics. Four AVT providers were included in the pilot and in one of the departments two solutions were compared in sequence.

The aim was to gain a rich, real-world understanding of how AVT could be implemented across the Trust in diverse settings, rather than limiting the approach to a tightly controlled research study. The pilot was guided by four principal aims:

to enhance clinical efficiency and workflow	to improve patient care
to support staff experience	to deliver cost savings

The evaluation used a combination of self-reported quantitative data, such as estimates of time saved on documentation, and qualitative insights from surveys, interviews, and user logs, all co-developed with departmental champions and AVT providers.

1.2 Summary of Key Findings

The AVT tools were generally well received by clinicians, with most reporting that the technology enabled them to complete documentation more quickly and efficiently. Nearly nine in ten users (88%) experienced measurable time savings, and many described a transformation in their working day, as administrative tasks that previously extended into evenings and weekends were now routinely completed within clinic hours. The most common benefit was a saving of 1-10 minutes per encounter (47.3%), with 45.5% saving 5-15 minutes and 22.2% of users reporting saving one hour or more per day. The mean time spent on documentation outside normal hours dropped by nearly half (from 3.02 to 1.55 hours/week), and the median fell from 2.0 to 0.75 hours/week. The percentage of users spending 5-10 hours of their own time on documentation dropped sharply from 24.2% pre-pilot to just 2.2% post-pilot.

This shift led to a reported improvement in work-life balance, reduced cognitive fatigue, and less stress, with staff repeatedly highlighting the positive impact on their wellbeing of these factors being reduced. Expectations of wellbeing improvements were high before the project, with 66% expecting AVT to improve their experience; these expectations were exceeded after real-world use, with 73% agreeing or strongly agreeing that they had experienced an improvement in wellbeing. The share of users indicating reduced stress rose by 70.2%, and reported work-life balance improved by 16.4%.
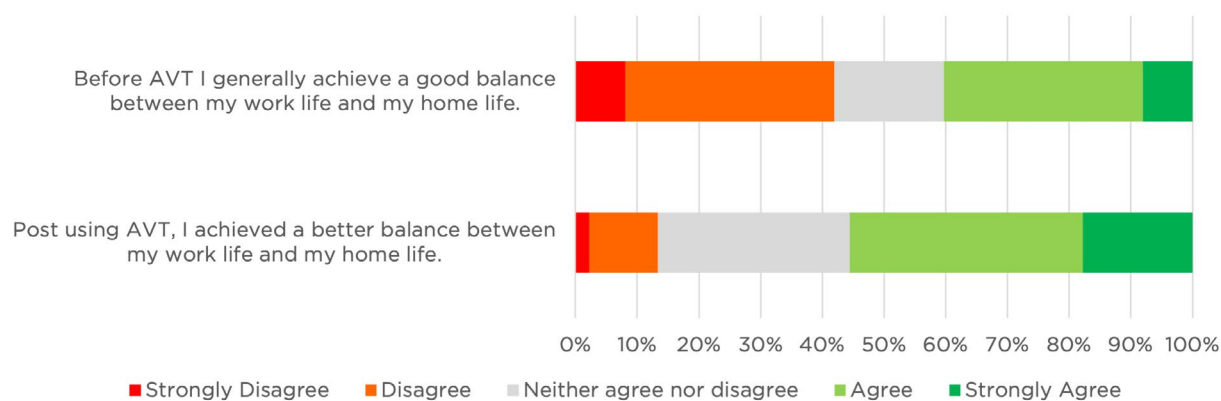
Figure 1: Representation of work-life balance.

A wellbeing metric from the NHS National Staff Survey was included in the pre and post-pilot survey: Before the project, 82% of all participants agreed or strongly agreed that they often felt worn out at the end of their shift or working day. After using AVT, more than half of all participants (57%) agreed or strongly agreed that they felt worn out at the end of their shift or working day less often.

Patient feedback, captured via staff responses, was overwhelmingly positive, with individuals noting that consultations felt more focused and that documentation was thorough and timely.

The pilot also revealed important limitations that should be considered in interpreting the results. While AVT tools reliably produced the bulk of the documentation, significantly reducing the effort required to create clinical records from scratch, almost all outputs required some degree of editing. Users routinely needed to review and amend details, such as correcting minor transcription errors, clarifying terminology, or removing occasional inaccuracies, before the documentation could be signed off.

In practice, 37.3% of logged AVT uses required the user to edit outputs for accuracy (126 out of 398 uses for accuracy of clinical information, 43 for spelling/punctuation/grammar, and 24 for invented/hallucinated information). When giving summative feedback in post-pilot surveys, hallucinations had been experienced by 44.4% of users in one or more of their encounters across the range of providers and use cases – illustrating that checking of notes by clinicians is an important process step.

The 24 usages where hallucination was identified as a reason for editing represent 6.2% of logged usages during the pilot. The descriptions of these hallucinations included: "added advice which was not given," "invented someone else in the room," "confabulated dates" and "made up growth measurements and centiles." These more significant hallucinations were most commonly observed in complex, multi-voice, or less structured clinical settings, where the technology struggled to capture the nuance and detail required for high-quality clinical records. A caveat to add here is that engagement with users during the pilot suggested that some were unclear as to the distinction between errors (caused by e.g. AVT mis-transcribing a word or a templating issue) and true hallucinations. This suggests that the 6.2% statistic may be over representative.

There was very little variation in accuracy between solutions. Rather, as explored below, the nature of the clinical workflow was the key determiner of accuracy.

For the purpose of the pilot, users were required to cut and paste AVT outputs between the AVT solution and other systems (e.g. electronic patient records). Most users experienced enough benefit from the standalone version of AVT that the inconvenience of this step was considered acceptable, given the overall reduction in documentation workload and the improvement in workflow. Nonetheless, the hope for integration with existing systems in the long term emerged as a persistent theme, with clinicians expressing minor frustration at the need to manually transfer outputs between systems.

In settings where AVT was well matched to the workflow, the benefits including wellbeing improvement were most pronounced, with high user satisfaction and minimal editing required. In Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM) and Ear Nose and Throat (ENT), for example, 85-90% of users reported that AVT saved them time, typically five to fifteen minutes per encounter, with some estimating up to thirty minutes saved in more complex clinics. Editing requirements were generally low, with 80% of users in these departments needing only minor changes to AVT-generated documents. User satisfaction was high: 92% agreed or strongly agreed that AVT was easy to use, and 88% felt it enabled them to focus more on patients and reduced after-hours admin.

In contrast, environments characterised by complex, qualitative documentation or frequent multidisciplinary input saw more mixed results, with some users finding that the time saved was offset by the need for extensive revision. For example, in Community Paediatrics, 60% of users reported that AVT required "a lot" of editing, and 40% reported frequent confabulation or omission of key details. Even in these cases, users emphasised that the AVT system still provided a valuable starting point, with the core content in place, and that the time and cognitive effort saved outweighed the inconvenience of editing.

1.3 Recommendations

It is recommended that AVT provision is strongly considered for reasons of staff wellbeing and overall productivity, with business cases in different organisations likely to differ in their strength and deliverability. A business case built solely on cash-releasing productivity benefits is likely to fail without careful consideration of implementation, but more nuanced business cases and a strong implementation plan is likely to lead to benefits for staff and for the organisation. The evaluation demonstrates that AVT consistently enhances clinician wellbeing and improves general productivity, with many staff reporting a substantial reduction in administrative burden, less stress, and a greater ability to complete documentation within working hours. In addition, while the primary gains relate to workflow and wellbeing, there is also evidence that, for a minority of clinicians and settings, AVT can create opportunities to increase capacity by enabling more patients to be seen or by allowing staff to manage their workload more efficiently.

To maximise these benefits, it remains important to balance the drive for standardisation in documentation with the flexibility to tailor templates to local needs, ensuring that the technology is closely aligned with the clinical context. Clear and accessible information about data governance is also essential to build and maintain trust among clinicians, many of whom remain uncertain about how AVT data are stored, accessed, and retained. A phased approach to deployment is advised, beginning with low-complexity environments where AVT has already demonstrated reliable performance, and using these as proof-of-concept sites to refine templates and workflows before expanding into more challenging areas. Any AVT implementation plan should take into account the broad range of benefits seen in the pilot and should prioritise licence allocation to departments that are confident they will realise a spectrum of benefits across our key themes, including efficiency, wellbeing, and quality of documentation.

Ultimately, the case for AVT rests on its proven ability to support clinician wellbeing and productivity, and while its impact on patient throughput and direct cost savings may be limited by broader organisational factors, its potential to improve the sustainability and quality of clinical work is clear, provided that implementation is carefully managed and responsive to the realities of frontline practice.

2. Introduction

2.1 Background

Ambient Voice Technology (AVT) is rapidly emerging as a potentially transformative tool in clinical settings, enabling handsfree interaction, real-time documentation, and improved workflow efficiency. Early trials and pilots suggested AVT could significantly reduce clinician administrative burden, shorten paperwork time related to patient contacts, and allow healthcare professionals to spend more time focused on patients. However, current adoption remains limited and cautious: oversight from bodies such as NHS England underscores that only AVT solutions meeting strict clinical safety, data protection and regulatory standards should be used, to safeguard patient confidentiality and clinical accuracy.

Oxford University Hospital (OUH) initiated an AVT pilot in July 2025 to explore the potential benefits of AI-enabled solutions for clinical documentation for patients and staff. The project emerged in response to increasing requests from staff for solutions that could reduce administrative burden and improve patient care, and the risk that some staff might be using one of the many AVT solutions available on individual licences in their private practice, with little understanding in the variability of these products. AVT tools have the ability to record, transcribe, and generate structured documentation, aiming to streamline workflows and enhance accuracy.

The pilot involved multiple departments, which were selected after broad engagement to capture data for a range of use cases and users, including high and low complexity settings, inpatient and outpatient services and consultants and nursing staff (see table on next page). The pilot was designed as a service evaluation rather than a research study, with findings intended to inform any potential procurement process and future implementation strategies.

The table below provides a breakdown of statistics of each solution throughout the pilot:

	Solution 1	Solution 2	Solution 3	Solution 4	TOTAL
Total number of OUH users during pilot	12	30	49	69	160*
Total number of encounters/dictations captured	182	1,479	3,747	3,831	9,239
Total number of minutes of transcription	2,294	27,430	75,471	94,456	199,651 days= 138 hours= 3327
Total number of outputs produced	331	1,759	2,418	2,027	6,535

*160 includes duplicates who used more than one solution. There were 136 unique users.

Section	Use Case/Workflow	Users	Complexity	Notes & Benefits
Audiology	Paediatric and adult appointments (history, diagnostics, counselling/ therapy) using AVT to listen, summarise, and generate reports/letters. Templates iterated locally; admin often completed during clinic with minor post-appointment clarifications.	Audiologists, clinical scientists	Low-Moderate	Standard clinics: easy to use, minimal editing (tympanometry terms, names), time savings 10–30 mins/day. Chaotic sessions: template optimisation needed; immediate check recommended.
Community Paediatrics	Neurodevelopmental assessments and reviews; AVT used for clinic letters and summary notes. Pre-clinic prep via AVT notepad; integration of external reports still manual.	Consultants and doctors	High	Benefits: reduced cognitive load, formatted starting point; typical savings 15–30 mins/day, 2–3 hrs/week. Constraints: frequent editing for tone, hallucinations (invented details), variable efficiency gains.
Oncology Pre-hyster Telephone Clinic	Structured telephone consultations for pre-hysteroscopy; AVT used to produce notes/letters matching EPR layouts.	Specialist nurses	Low	Editing and reformatting offset time savings; net gain ~1–2 mins per patient.
Critical Care MDT Ward Round	Ambient capture of multi-voice ward-round discussions and plans; some consultants tested AVT in clinics.	Consultants and resident doctors	High	Mixed outcomes: SHO-level summaries, duplication, hallucinations; integration essential. Best suited for family discussions or end-of-round summaries.
Critical Care Inpatient Nursing	ICU nurses used AVT for dictating admission/ start-of-shift assessments; summarised in systems-based templates.	Critical care nurses	Moderate	Consistent time savings (10–30 mins per encounter); minor edits for hallucinations and structure fixes. Integration and hardware issues noted; dictation-first approach preferred.
OCDEM Consultants	Diabetes/endocrine outpatient consultations and acute medicine clinics; AVT generated letters and summaries.	Consultants	Low-Moderate	Strong positive feedback; admin savings 15–60 mins/day, ≤2–3 hrs/week; minor edits for tone and drug names. Integration and template personalisation requested.
OCDEM Diabetes Outpatient and Inpatient Nursing	Specialist nurses used AVT in clinics, advice lines, and antenatal care; generated structured letters and contemporaneous notes.	Diabetes specialist nurses	Low-Moderate	Time savings (15–60 mins/day); improved work-life balance; minor edits (insulin names). Advice lines: enabled same-day sign-off and doubled throughput; integration needed.

Use cases ongoing at time of report: Dermatology and Inpatient Psychiatry.

2.2 Project Aims

The AVT pilot was structured around four primary research aims, each with specific secondary questions. These questions covered key value domains that would indicate the likely impact and benefit of procuring and deploying AVT in a range of settings. We also looked to understand whether existing digital dictation provision could be consolidated within AVT provision.

Does AVT enhance clinical efficiency and workflow?

- Does AVT reduce time spent on clinical notes and letters?
- Does AVT enable more timely management decisions and actions?
- Could AVT reduce waiting lists or accelerate patient flow?

Does AVT improve patient care?

- Does AVT increase face-to-face time between clinicians and patients?
- Does AVT create more effective clinical documentation?
- Does AVT capture more clinically standardised, accurate, and actionable data?

Does AVT improve staff experience?

- Does AVT improve staff experience of patient encounters and clinical conversations?
- Does AVT improve work-life balance for staff?
- Does AVT function effectively across a representative range of OUH use cases?

Does AVT deliver cost savings?

- Could operational efficiencies from AVT be converted into cash-releasing changes?
- Could improvements in work-life balance reduce staff absence and turnover?

These aims and questions guide both the design and evaluation of the AVT pilot, ensuring that outcomes are measured across efficiency, care quality, staff wellbeing, and financial impact.

2.3 Project Design

Prior to project initiation, three online engagement sessions were held to gather input and expressions of interest from colleagues and departments across the Trust. These sessions were supported by a targeted email campaign to key individuals within each clinical directorate, seeking information on current AVT usage, priority criteria for selecting an OUH-provided AVT solution, and willingness to participate as a pilot site.

Attendance and interest at the engagement sessions was high and participating departments and clinical champions were recruited from among the responders. The project was then co-developed with clinical champions alongside representatives from the selected providers. Indicative metrics covering a deliberately wide range of variables were favoured over highly controlled and standardised datasets.

The pilot was structured in four phases:

Phase 0
A range of providers were interviewed and assessed for potential inclusion in the pilot.

Phase 1
TheHill's team worked with companies and champions in the services to iterate local templates and ensure the tools were ready for deployment in those areas.

Phase 2
Focused on feasibility, running for six weeks across ENT, Audiology, OCDEM, Oncology and Critical Care. Tools were tested as standalone web applications without integration into EPR systems, requiring manual transfer of documentation.

Phase 3
Expanded to Community Paediatrics.

Four AVT solutions were used for the evaluation. The initial longlist of solutions was informed by structured market scanning, including a review of products already in use within the NHS, emerging solutions demonstrating clinical maturity, and technologies referenced through peer Trusts and national networks. Particular consideration was given to solution readiness for deployment in an acute setting, evidence of clinical use at scale, and alignment with NHS information governance and digital safety requirements. From this process, four solutions were shortlisted for further evaluation to provide a balanced view of the market and reduce dependency on a single vendor archetype.

To achieve this balance, two shortlisted solutions were from established healthcare technology providers with existing NHS footprints, and two from vendors specialising specifically in ambient voice and AI-enabled clinical documentation. During initial validation and onboarding, one solution from each category did not progress due to failing to meet early technical, operational, or governance requirements. These were replaced with two further solutions: one that had already gained organic adoption within the Trust through privately procured licences, and another that had demonstrated traction across both primary and secondary care settings. This approach ensured that the final four solutions represented a pragmatic mix of market maturity, specialist capability, and real-world clinical adoption, enabling a robust and comparative evaluation during the pilot phase.

The solutions eventually tested were Accurx Scribe, Heidi, Lyrebird and TORTUS, with one of these being deployed in more than one use case. This was both to compare the outcomes for individual solutions between different use cases and also to identify variation in the performance of different solutions when deployed in the same setting.

Each solution was deployed on Trust devices where possible, with personal devices permitted under governance controls. For the purposes of the pilot, outputs were manually copied from a web browser version of each solution into Trust digital records.

All participating staff completed information governance and AI training prior to onboarding.

To capture data, users completed a pre-pilot survey, multiple user logs of discrete AVT usages, and a post-pilot survey. This provided three comparable datasets: a high-level baseline of expectations and current experience, granular logging of AVT performance in individual patient encounters and finally a high-level review of user experience at the end of the project. Selected users then participated in semi-structured interviews. Data was analysed statistically and thematically. MS Co-pilot was used where appropriate to support analytical and summarisation tasks.

Solutions deployed across OUH during pilot:

3. Key findings in detail

3.1 Accuracy

In use, the AVT scored well for accuracy. Users completing the in-pilot data capture agreed or strongly agreed for 81.5% of logged encounters that the AVT "produced an accurate record of the encounters and its agreed outcomes." Users agreed or strongly agreed for 83% of logged encounters that "the AVT accurately captured concepts and terminology." When asked to review accuracy overall in post-pilot surveys, 40% of respondents agreed or strongly agreed that AVT-produced documentation was "more clinically accurate than normal practice."

37.3% of the sample AVT usages logged in the in-pilot data capture required the user to edit outputs for accuracy. Out of 398 logged encounters overall, 126 required editing for accuracy of clinical information, 43 for accuracy of spelling, punctuation and grammar and 24 for invented (hallucinated) information. Free text entries reported specific mis-hearings, number capture errors, drug names mistranscribed, and hallucinations (e.g. invented details or wrong attributions).

The high frequency of inaccuracies logged as reasons for editing contrasts with the overall positive sentiment about AVT accuracy, indicating that most errors were minor and did not diminish users' favourable views. Post-pilot surveys and interviews confirm this, with some respondents even noting that minor inaccuracies served as helpful prompts to carefully proofread all outputs.

In post pilot surveys, 44.4% of project participants agreed that "I encountered situations where the AV system/s invented (hallucinated) information". This indicates that hallucinations were not confined t specific users but occurred for a wide range of project participants. However, engagement with user during the pilot suggested that some were unclear as to the distinction between errors (caused by e.g. AV mistranscribing a word or a templating issue) and true hallucinations.

Across all datasets, some key trends were identified:

- Specialties with more structured, standardised, and one-to-one encounters (e.g., OCDEM clinics, Audiology after template optimisation) experienced fewer and less severe hallucinations.

- Specialties with complex, qualitative, or multi-voice encounters (e.g., Community Paediatrics, Critical Care ward rounds) experienced more frequent and/or severe hallucinations.

- Hallucinations differed by specialty: factual errors in OCDEM, narrative confabulation in Critical Care, template-driven assumptions in Audiology, and unsupported conclusions in Community Paediatrics. Testing of more than one solution in one of these use cases indicates that hallucination rate correlated with use cases rather than the AVT solution used.

- Some AVT solutions (notably one solution tested in Community Paediatrics) were more prone to hallucinations in complex settings, but template design and context were also critical factors, with the same solution experiencing very few hallucinations in the more bounded, standardised OCDEM use cases.

3.2 Trust

Participants' trust in the solution was tested using the pre and post pilot surveys, assessed using the Mayer, Davis and Schoorman (1995) framework.

For pre-pilot trust, clinicians were hopeful about the potential benefits of AVT but reserved in their confidence. Ability-based trust (confidence in accurate transcription) was the weakest dimension (60% agreement; mean = 3.62/5). Free-text comments frequently mentioned worries about transcription errors, hallucinations, and whether the systems could cope with real clinical complexity. Integrity-based trust was slightly higher (68% agreement; mean = 3.83/5), likely influenced by confidence in NHS data governance, though in the free-text comments some clinicians wanted clearer explanations of how data would be stored and used. Benevolence-based trust was the strongest pre-pilot dimension (64% agreement; mean = 3.89/5), with many clinicians believing AVT tools were designed with their needs centered, while also expressing a desire for collaboration between developers and clinicians.

Post-pilot trust increased across all three dimensions. Ability-based trust increased from 60% to 70% agreement (mean = 4.11/5), with many clinicians describing the outputs as accurate and reliable except in especially complex cases. Integrity-based trust showed only a very small change (66%; mean = 3.98/5), reflecting that hands-on use did not provide additional insight into data flows and in free-text comments, several clinicians continued to request clearer information on storage and deletion. Benevolence-based trust increased the most (from 64% to 75%; mean = 4.14/5). Clinicians reported reduced cognitive load, improved patient interaction and faster documentation, leading them to view the tools as better aligned with real clinical needs.

3.3 Accuracy of AVT as a dictation tool

Evidence from in-pilot user logs shows that AVT performed well in direct dictation tasks. Among the 92 entries where users rated the quality of dictation, 69.6% described it as either excellent or good, 27.2% found it reasonable, and only 2.2% rated it poor.
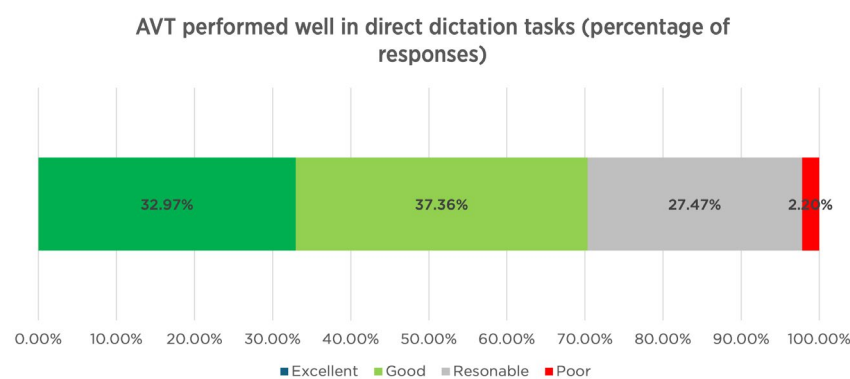

AVT performed well in direct dictation tasks (percentage of responses)

| 32.97% | 37.36% | 27.47% | 2.20% |

■ Excellent ■ Good ■ Resonable ■ Poor

Figure 2: Performance of AVT in direct dictation tasks

Qualitative feedback from interviews and comments reinforces this picture. Clinicians who used AVT for straightforward dictation frequently reported that it was easy to use, required only minor edits, and handled accents far better than the Trust's main dictation tool. Several users stated that the current solution struggled with their accents and often required them to retype entire letters, whereas AVT produced usable drafts with fewer errors. Some corrections were still needed, particularly for numbers and medication names, but these were described as occasional rather than routine.

No quantitative ratings for the Trust's main tool are available in the pilot data, but the qualitative evidence points to AVT being a clear improvement on existing options for dictation tasks in this pilot.

3.4 Time Saving and Efficiency

Pre-Pilot

Time savings were anticipated and hoped for by many participants ahead of the pilot. In pre pilot surveys, 88.7% agreed or strongly agreed that AVT can save time on admin tasks for them. Free-text comments expressed a hope that AVT would help to reduce a sometimes overwhelming burden of documentation: "Maybe this AVT thingy will help me get back on top of things and allow me to provide a higher standard of care by completing [administrative] tasks contemporaneously."

The post pilot interviews also supported this with comments such as: "Anything that helps us with our admin loads, I consider a positive thing. [Documentation] takes a really, really long time. [It] very soon builds up and impacts on clinics."

88.7%

of pre-pilot survey respondents agreed or strongly agreed that AVT can save time on admin tasks.
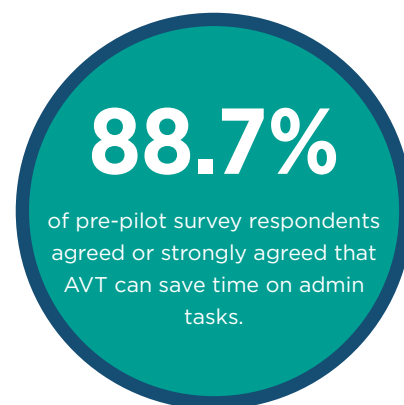
During and after pilot

Encounters logged on the in-pilot data capture show that AVT achieved substantial time savings for the vast majority of users, with nearly 9 in 10 (86.9%) reporting a measurable reduction in documentation time. The most common benefit was a saving of 1–10 minutes per encounter (47.3%) with 45.5% saving 5-15 minutes. A significant minority reported even greater efficiencies including 4.4% achieving a >30 minute time saving on a single task.
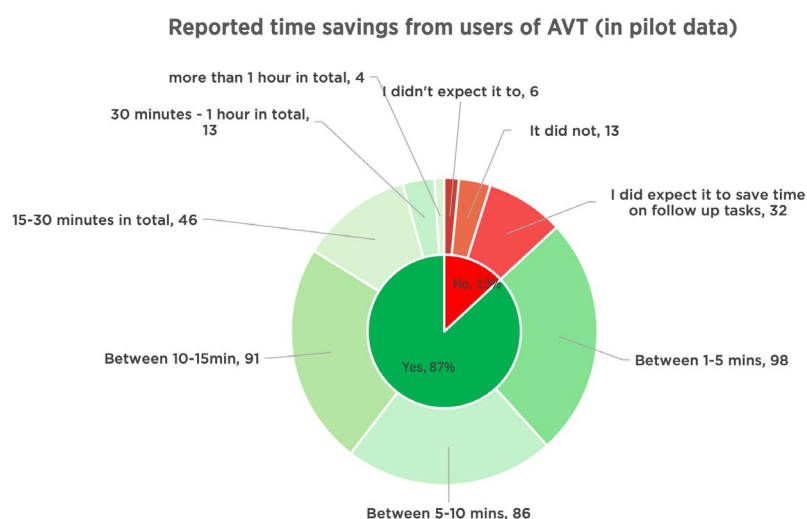

Reported time savings from users of AVT (in pilot data)

- more than 1 hour in total, 4
- I didn't expect it to, 6
- 30 minutes - 1 hour in total, 13
- It did not, 13
- 15-30 minutes in total, 46
- I did expect it to save time on follow up tasks, 32
- No, 13%
- Between 10-15min, 91
- Yes, 87%
- Between 1-5 mins, 98
- Between 5-10 mins, 86

Figure 3: Did the AVT solution enable you to complete tasks quicker for this encounter than normal practice?

When asked to reflect on perceived reduction in documentation time in post-pilot surveys, over half of users (55.6%) reported saving at least 30 minutes per day, including 1 in 5 (22.2%) who reported saving 1 hour or more per day. Very few (5%) reported no or negative time savings. The median daily time saving was approximately 45 minutes per day, as reported by users.

Post-pilot survey data show that overall 77.8% of respondents agreed or strongly agreed that AVT saved them time on administrative tasks, and 82.2% agreed or strongly agreed that it increased face-to-face time and reduced computer use during patient encounters. This strong signal for improved in-clinic workflow was echoed in qualitative feedback.

However, the benefit to departmental admin time was less pronounced: only 51.2% agreed or strongly agreed that AVT saved time for others in their department, with a large proportion of neutral responses, suggesting that wider team benefits may depend on local workflows and roles.

The majority of respondents do not expect AVT systems to directly increase patient flow or result in more patients being seen per clinic session; only 24.4% agreed or strongly agreed that AVT helped them see more patients or increase income, with many neutral or disagreeing. This suggests that, while AVT delivers clear efficiency gains for documentation and in-clinic workflow, these do not automatically translate into increased throughput, likely due to scheduling and integration constraints, as well as unpaid hours being done for some clinicians in the current model.

A pilot area where the structure of the service allowed increased throughput was the advice line staffed by Diabetes Specialist Nurses. Here, AVT users consistently fed back that the tools dramatically increased the number of unscheduled calls they were able to answer and complete documentation for. One nurse shared that they were able to "take twice as many calls on the advice line" without pausing to type notes.

Qualitative feedback captured during in-pilot usage logging, post-pilot surveys, and participant interviews reinforces these findings. Across a wide range of clinical settings, users consistently reported that AVT systems delivered substantial efficiency gains. In structured outpatient clinics, the impact was especially pronounced, with users describing AVT as a "game changer" for reducing admin burden and enabling more patient-centred care.

While a small minority noted that editing or technical issues could occasionally delay the process, the overall consensus was that AVT significantly shortens the time-to-letter for patients, supporting more efficient and responsive care. The resources available did not allow us to measure this directly, but auditing the time from clinical to letter production would be a useful next step.

Some users did experience technical or usability issues, such as occasional system downtime, formatting challenges, or the need for careful proofreading, particularly in complex or noisy environments. A minority found that frequent editing or inconsistent performance could offset some of the time saved, especially in highly complex or multidisciplinary settings.

A significant amount of the time saved with AVT appears to have been reclaimed from unpaid, non-contracted work. When comparing the normal amount of time spent on documentation per week pre-pilot and while using AVT, the mean time spent on documentation outside normal hours dropped by nearly half (from 3.02 to 1.55 hours/week) after AVT was introduced, and the median fell from 2.0 to 0.75 hours/week. This indicates that a typical user now spends less than an hour per week outside normal hours on documentation.
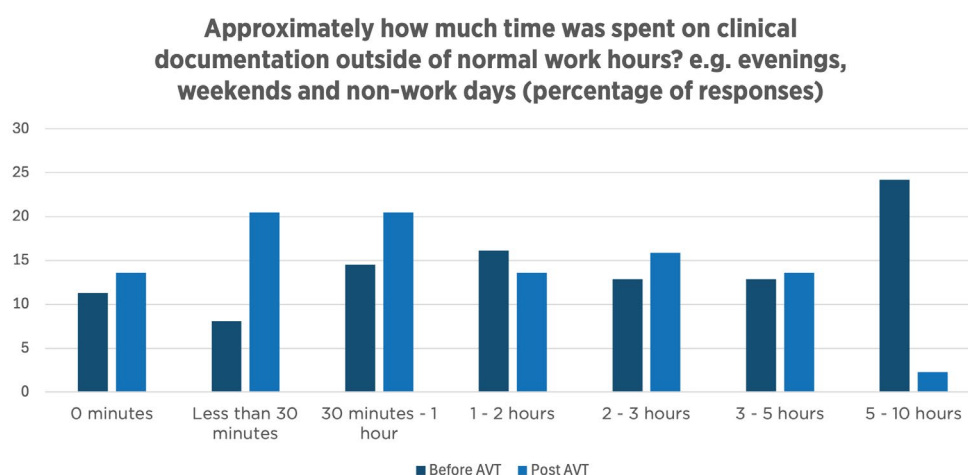

Figure 4: Time spent on clinical documentation outside of normal working hours (all respondents)

The improvement was particularly marked for users who reported spending 5–10 hours of their own time on documentation per week. When examining a subset of data where responses could be matched to individual named participants pre pilot and post pilot (rather than the full datasets, where not all participants completed both surveys), there was a reduction from 26.7% per week spending 5-10 hours to 0% spending this amount of time. Most users now report spending less than 2 hours per week (post-pilot: 66.7% in categories up to "0–2 hours", compared to 50% pre-pilot). Several noted that tasks which previously required evening or weekend work could now be finished during the working day. One consultant shared: "I used to leave at 8pm regularly; now my letters are usually done by 6pm"

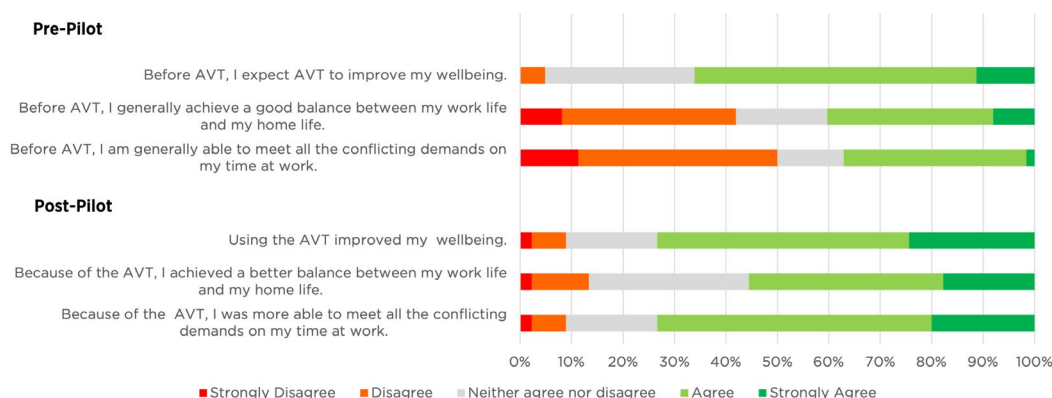## 3.5 Staff Experience and Wellbeing

Figure 5: Experience and wellbeing before pre and post pilot.

Expectations of wellbeing improvements were high before the project with 66.1% expecting AVT to improve their experience in this regard. These expectations were exceeded after real-world use with 73.3% agreeing or strongly agreeing that they had experienced an improvement in wellbeing. Qualitative comments describe less exhaustion, fewer headaches, finishing admin within clinic, and not working weekends to complete letters.

Many staff described a tangible reduction in mental fatigue and "cognitive burden". For example, one nurse said:

> *I used to have brain fog in the afternoon, but now I have more energy because I'm not focusing so much on language. Definitely less drained, because the focus on language was lifted.*
>
> *Specialist Nurse*

Several clinicians highlighted that AVT allowed them to finish admin within their working hours, rather than taking work home or working weekends:

> *Now, unless it's an exceptionally busy day, I can complete about seven out of ten letters immediately after consultations. Previously, I'd only start on them the next day and often had to use a weekend day to catch up, which I no longer need to do.*
>
> *Specialist Nurse*

> *Not having the burden of extra paperwork has improved my wellbeing and job satisfaction. Losing AVT would feel like a step backwards.*
>
> *Specialist Nurse*

The benefits were experienced in a number of ways: After using AVT, the share of respondents who felt more able to meet conflicting demands increased by 35.7 percentage points. Reported work–life balance improved by 16.4 percentage points.

The ability to finish documentation within clinic hours was described as "transformational" for work–life balance:

> " *My main saving is my personal time. During consultations, I focus on the patient and take brief notes, rather than typing constantly. AVT helps me recall everything that was said and make sense of it, especially with patients who talk at length. Now I can finish my admin in a timely manner, and it doesn't pile up.* "

Specialist Nurse

> " *After a back-to-back clinic, all admin would be done by the end of the day, which was previously unheard of. Reports went out on the day, families received them quicker, and referrals happen faster. It made a huge difference.* "

Audiologist

Staff repeatedly described improved experience from being able to focus more on patients during consultations, rather than splitting attention between the patient and the computer:

> " *AVT let me focus on patients without taking notes, which enhanced the quality of interactions. Patients responded positively, appreciating the more open conversation. The system picked up details well, even with softly spoken individuals or those for whom English isn't their first language. It reduced my stress, knowing I wouldn't forget important points.* "

Specialist Nurse

In pre-pilot surveys, 71% of respondents agreed or strongly agreed that they "often experience stress at work". After using AVT a majority of respondents said they had seen an improvement in this area and agreed or strongly agreed that they experienced stress less often (73.3%).

A question was also asked that aligns with NHS National Staff Survey wording: "I often feel worn out at the end of my working day/shift." Before the project, 82.3% of participants agreed or strongly agreed that they often felt worn out at the end of their working day. After using AVT, more than half of participants (55.6%) agreed or strongly agreed that this measure had improved for them.

Many clinicians described a marked reduction in stress and exhaustion:

> " *I feel more positive. Before, if I had a full clinic, I'd wonder how I'd write all the letters. Now, even if it's a busy day, it's OK.* "
>
> *Specialist Nurse*

> " *I used to leave at 8pm regularly; now my letters are usually done by 6pm. I've even managed a 13-patient clinic, [which would normally be 10 patients] knowing I can get the letters drafted.* "
>
> *Consultant*

73.3%

of respondents said they had seen an improvement in stress levels and experienced stress less often.

> " *It's less mentally exhausting at the end of a long day. AVT takes some of the pressure off, allowing me to check the notes without having to think so hard in English. It's a life changer.* "
>
> *Specialist Nurse*

> " *The main benefit is reducing stress. After a clinic, if I can't do the administration and report writing soon after, it becomes much harder to remember details. With AVT, I had all my clinic letters done quickly, which transformed a potentially overwhelming week into something manageable.* "
>
> *Consultant*

3.6 Patient Experience

Of 395 encounters logged in the in-pilot data capture sample, only 1 patient declined to have AVT used, equating to 0.3%. Explicit patient feedback about AVT, as reported by staff in both the post-pilot survey and interview transcripts, was consistently positive and provides a clear picture of how patients themselves experienced the technology.

Direct comments from patients were recorded in 22.7% of user survey responses, with individuals describing the process as "efficient", "thorough", and "helpful". Several clinicians recounted patients expressing interest or curiosity about the use of AI, and some asked to see the AI-generated letter or transcript; all who did so expressed satisfaction with the content and accuracy. One patient specifically compared their experience to another hospital, noting that "a previous doctor didn't look at me, but with AVT, you looked at me throughout the consultation", highlighting a perceived improvement in the quality of the clinical interaction.

In post-pilot feedback from Audiology colleagues, every clinician commented that families were positive or supportive of AVT use, with one parent describing AI as "a bit evil" but still consenting and expressing interest in the technology. Parents in paediatrics occasionally asked about the process, and while one expressed concern about NHS staff being "replaced", they ultimately accepted the use of AVT.

Patients also commented on the documentation itself. Clinicians observed patients stating that the documentation "captured everything they wanted" and that patients were "really impressed" at how it reflected their story. In specialties such as audiology and ENT, colleagues noted that reports were more detailed and "arrived quicker".

It was not possible to directly compare time-to-letter with AVT against standard practice, but qualitative feedback indicates that AVT shortened this interval, leading to patients and GPs receiving letters more promptly. Survey responses and interviews indicated a clear time reduction, with many users reporting that letters and reports were now routinely finished within clinic hours rather than being delayed until later in the day or week. Interviewees across several departments described being able to send out letters immediately after consultations, attributing this improvement directly to AVT's ability to generate a first draft in real time. This acceleration of documentation workflows was particularly evident in OCDEM and ENT, but was also noted in other areas.

No explicit negative feedback from patients about the AVT-assisted experience was reported by staff in either dataset. The only concerns raised were about the novelty of AI or curiosity about its use, not about the impact on their experience.

The introduction of Ambient Voice Technology (AVT) had a marked effect on clinicians' ability to be present during patient encounters. According to the post-pilot user survey, a striking 81.8% of clinicians agreed or strongly agreed that AVT increased face-to-face time and reduced computer use during clinical encounters.

81.8%
of clinicians agreed or strongly agreed that AVT increased face-to-face time and reduced computer use during clinical encounters.

Clinicians consistently described how AVT allowed them to focus more fully on patients, rather than splitting attention between the patient and the computer.

Resources did not allow direct tracking of patient feedback but the 2nd hand data that was captured via clinical staff indicates a likely improvement in patient experience as a result of AVT usage, and the technology was well accepted by patients. Further work could include direct patient feedback.

3.7 Variation Between Departments / Use Cases

In departments with large numbers of users, analysis of AVT performance demonstrates clear and statistically significant differences, with OCDEM and ENT achieving the highest effectiveness and Community Paediatrics and Critical Care consistently reporting lower ratings.

These differences are strongly linked to workflow characteristics: AVT was most effective in settings with standardised, linear processes and adaptable templates, while its performance declined in environments where documentation was complex, multidisciplinary, or required nuanced interpretation. The in-pilot data capture provides robust quantitative evidence for these patterns. In OCDEM and ENT, 83.9% of users reported that AVT saved them time. The average time saved per encounter among those who reported a benefit was 11.6 minutes. When averaged across all encounters (including those who did not report a benefit), the time saved was 10.2 minutes per encounter.

9.9mins

time saved per encounter, averaged across all encounters (inc. those who did not report a benefit).

Qualitative feedback echoed these findings, with clinicians describing improved workflow, reduced and the ability to complete documentation within clinic hours. AVT's handling of accents and back noise was also positively noted, and even less tech-confident users found it accessible.

By contrast, Community Paediatrics and Critical Care, with much more complex workflows and ten faced persistent challenges. Post-pilot interviews revealed a frequent need to correct errors in AVT output. In Critical Care, one consultant described the transcript as "not accurate enough for our needs". Qualitative feedback from these areas highlighted AVT's tendency to summarise or invent information, misattribute statements, and require repeated prompting, with the editing burden sometimes outweighing any time saved. Solutions that performed well in less complex areas within the same pilot saw an increased error rate in more complex areas.

3.8 Variation Between Staff Groups

Analysis of both quantitative and qualitative data shows that specialist nurses, allied health professionals, and audiologists working in standard outpatient or clinic environments consistently reported high benefit. In the post-pilot survey, the mean effectiveness score for allied health professionals and audiologists was 5.0 (standard deviation 0.0), with 5 corresponding to a likert scale response of strongly agree, and for nurses it was 4.38 (standard deviation 0.52), 100% of these respondents agreeing or strongly agreeing that AVT was effective. One specialist nurse described AVT as "one of the most useful tools we've tried. I'd love to keep it," while an audiologist commented, "AVT allowed clinicians to have therapeutic conversations without worrying about missing information or breaking the flow to take notes." These groups highlighted time savings, reduced cognitive burden, and improved work-life balance as key benefits.

Consultants' experiences were more varied, with an overall mean effectiveness score of 4.05 (standard deviation 1.07) and 81% agreement. This apparent difference was largely attributable to the settings in which many consultants piloted AVT. In standardised, bounded use cases such as Endocrinology (within OCDEM), consultants experienced few issues, whereas in challenging environments such as Critical Care and Community Paediatrics, the complexity of multi-voice ward rounds and nuanced psychosocial consultations posed significant barriers.

As one Critical Care consultant reflected, "Within the current workflow, we spent a lot of time optimising the tool, which distracted from the ward round... with the tool, [ward rounds] actually took longer." In community paediatrics, another consultant noted that for qualitative, social, emotional, and behavioural information, AVT struggled: "It was all or nothing, and not reliable in recording key information."

Despite these challenges, the majority of consultants in Community Paediatrics, where the complexity of cases, consultations and documentation caused significant challenges, still reported significant benefits and expressed a desire to continue using AVT. 71% of Community Paediatrics consultants surveyed (and 4 out of the 5 interviewed post-pilot) indicated that AVT improved their workflow and reduced administrative burden, with one stating, "Overall, it's definitely been a benefit... I'm less likely to push myself late into the evening. It's helped me stay on top of things."

71%

of Community Paediatrics consultants surveyed indicated that AVT improved their workflow and reduced admin burden.

In summary, users of all professional backgrounds experienced benefits from AVT. Differences in experience were primarily explained by clinical setting and workflow demands, with the greatest challenges arising in complex, non-standard use cases rather than from the professional background of the user.
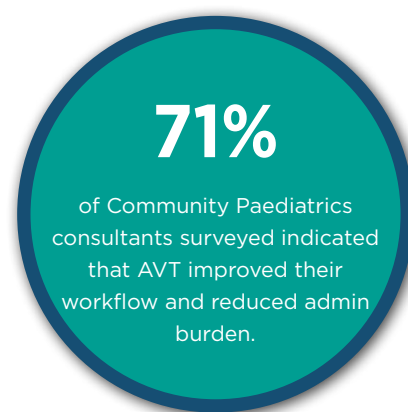
3.9 Variation Between Products and Providers

No single AVT solution emerged as universally superior during the pilot. Where individuals or teams expressed strong preferences, this appeared to be down to local factors or limitations of the pilot, rather than significant difference in solution performance. For example, one solution experienced connectivity failures that hadn't been seen in a different department on a different site. Another solution presented a formatting issue when pasting text into EPR that would in fact be resolved in a full (non-pilot) deployment by the provider coding highly customised paste formatting to the specific preference of a paying customer. In both these cases, the colleagues involved expressed lower approval of the less successful solution, but the result was not clearly a failing of the AVT solution itself.

Solutions trialled in multiple departments tended to show particular strengths in standard, single-clinician outpatient settings, especially for users with accents or working in a second language, delivering reliable accuracy and time savings. However, the data set was not large enough for robust statistical comparisons, and in more complex, multi-speaker, or highly qualitative environments (such as critical care, multidisciplinary clinics, and community paediatrics), all solutions faced limitations. User preferences in these settings often depended on factors like template flexibility, tone, and integration. Careful review and adaptation remain essential for any AVT deployment.

Similar challenges were reported across the dataset. 12% of users experienced system crashes, slow uploads, or lost transcripts, and 10% duplicated work as a precaution – likely due to connectivity issues rather than specific to one provider. Calls for improvement were common: In post pilot free-text comments, 46.7% of users requested better templates and customisation, 38.9% saw room for improved accuracy and quality, 20% identified opportunities for enhanced training and support, 13.3% requested the ability to include tables and other structured elements and 8.9% observed a need for improved terminology, tone and readability.

Two providers initially invited to participate in the pilot were ultimately excluded because their tools did not deliver the core functionality that had been expected based on initial discussions. This highlights the importance of rigorous user testing or evidence of other deployments for any shortlisted suppliers, as not all commercially available AVT solutions are truly ready for deployment in real-world clinical settings.

3.10 Challenges with Templates

AVT tools work by capturing a transcript of clinical conversations and then using a preset template to select salient information from the transcript and structure it into the required format, usually an EPR note and/or a patient letter.

Effective templates are essential as they guide the AI models behind the AVT technology to accurately select and organise key information. Ineffective templates can drive errors by e.g. prompting the solution to provide content that wasn't explicitly present in the transcript, causing the model to select inappropriate or irrelevant information.

Most solutions offer a selection of standard templates (e.g. to follow the Subjective, Objective, Assessment, and Plan (SOAP) structure) that are commonplace across healthcare settings. Hospital services have typically evolved bespoke variations on these conventional structures that better capture the specialised nature of their consultations, pathways and patient cohorts.

Pilot departments worked with providers to create tailored templates before going live with the AVT solutions. Nonetheless, template-related errors and inefficiencies were evident.

In the in-pilot logs of individual AVT uses, just 4.2% of respondents directly cited AVT template issues, while 46.7% did so in the post-pilot survey and 16.7% in interviews. These challenges were not universal, but they were significant for a subset of users, especially in clinical areas with complex documentation needs. Typical comments included "templates need improving," "still need to adapt into clinic template - editing," and "default templates made assumptions." The higher rates in surveys and interviews suggest that template problems became more apparent as users reflected on their experience and were prompted to discuss specific barriers.
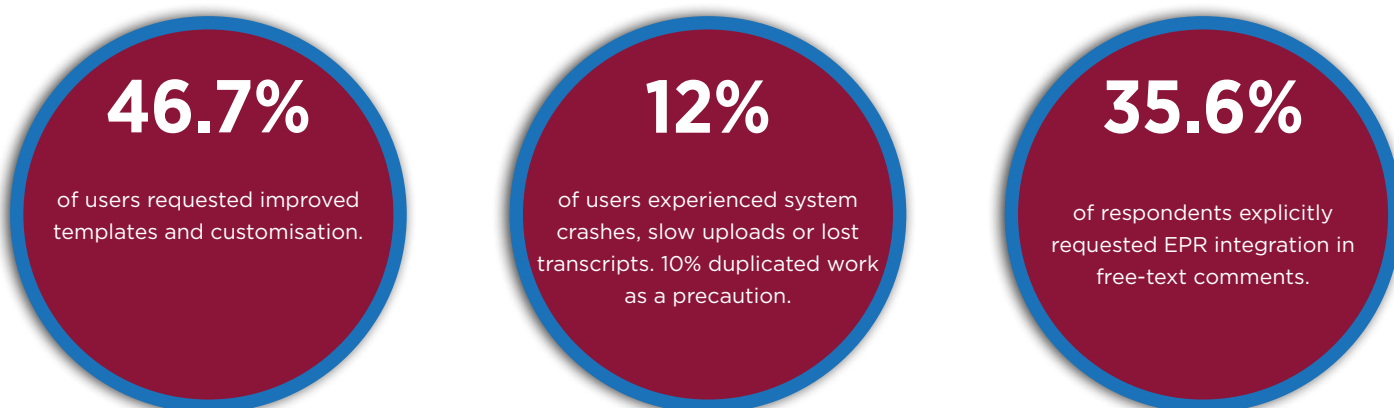
Users faced challenges such as poor alignment with local standards, missing or mis-ordered content, excessive editing, and unstable formatting. They addressed these by adapting AVT outputs to clinic templates, refining them over time, borrowing templates from colleagues, and setting rules to prevent assumptions. Of those reporting template issues, 33.3% in forms, 60.0% in the survey, and 19/24 in interviews described solutions like "edited to match clinic template" and "asked [the supplier] to make a template for medicine." These actions show local adaptation, iterative refinement, and collaboration.

Users who experienced template challenges made clear recommendations for future improvement. Among challenge reporters, 61.1% in the forms, 70.0% in the survey, and 14/24 in interviews offered explicit suggestions. These included calls for more customisable and role-specific templates, the ability to upload and edit templates, better export fidelity for integration with local systems, and standardisation across departments. Users also requested options to limit template length, suppress irrelevant sections, and introduce standard disclaimers for AI-generated content. Enhanced training and support, such as template repositories and "floor-walker" style help, were also recommended to help clinicians build and maintain effective templates.

Development of a template strategy will be essential to the success of AVT implementation.

3.11 Other Challenges

Similar challenges were reported across the dataset, across different providers and settings.

46.7%
of users requested improved templates and customisation.

12%
of users experienced system crashes, slow uploads or lost transcripts. 10% duplicated work as a precaution.

35.6%
of respondents explicitly requested EPR integration in free-text comments.

Many users described a learning curve, with persistent issues in handling complex or non-linear documentation, and many remarked that AVT's output often required "personalising" to match their usual clinical style. Describing how the tools' outputs improved with time and familiarity, users variously commented that "I found that with more use, I got better at prompting the system and got better results", "There is a learning curve, but after a few clinics it became much easier and the output improved" and "At first, I had to check everything, but as I got used to it and set up my templates, it needed less editing."

For most use cases, users wanted to be able to click a button and automatically input the note or the letter into EPR, rather than having to copy and paste. Being able to open the AVT within EPR would also be of benefit to these users to lower the complexity of workflow. More complex EPR integration (updating specific fields) was only deemed relevant in one use case (OCC). However there are use cases within the Trust that were not included in this pilot (e.g. Radiology reporting) where a more complex degree of integration would be needed to deliver the desired functionality.

35.6% of users explicitly requested integration in free-text comments, and the majority of users agreed that integration would be beneficial. However users were clear that the AVT solutions still delivered benefits without integration, and were reluctant to accept a lower quality of transcription in order to achieve EPR integration. None of the solutions tested in this pilot were integrated and all offered broadly the same level of service to one another.

These challenges show that AVT is still an emerging technology with plenty of room for improvement, and that there will be a learning curve for users as they adopt the technology.

3.12 Return on Investment

Across all three data sets (survey responses, free-text comments, and interview transcripts) there is strong and consistent evidence that users felt AVT improved their general productivity. The most frequently cited benefit was the reduction in time spent on clinical documentation, with many users reporting that letters and notes could be completed either contemporaneously or much more quickly after patient encounters. This time saving was described as "massive" by some, with several clinicians noting that they were able to finish their admin within working hours rather than staying late or working at weekends.

Users also described a significant reduction in cognitive and emotional burden: knowing that AVT would reliably capture the substance of a consultation meant less mental effort was required to recall details later, and many reported feeling less stressed and less fatigued at the end of the day. The ability to focus more fully on patients during appointments, rather than dividing attention between the conversation and note-taking, was repeatedly highlighted as a qualitative improvement in both workflow and patient experience. Even in cases where AVT did not directly increase patient throughput, users valued the opportunity to redirect time and energy towards other essential tasks, such as emails, leadership and line management responsibilities, service improvement, or teaching. This is likely to lead to a general improvement in quality and safety, although the benefits will be distributed and thus difficult to quantify.

Most users did not perceive AVT as a tool for increasing patient throughput or generating direct financial savings at the departmental level. Only a small minority reported being able to see more patients as a result of AVT, and this was typically limited to settings with flexible clinic templates where time savings could be directly converted into additional appointments. Many clinicians noted that, despite efficiency gains, organisational constraints such as fixed clinic schedules and pre-booked appointments meant that increased capacity was rarely realised in practice.

These findings highlight the importance of considering structural and organisational factors when implementing AVT. A department-specific approach to structural and process changes could yield direct patient throughput benefits if sympathetically designed and carefully thought through. However, the main value of AVT, as reflected in both survey and interview feedback, was seen in qualitative improvements, such as reduced stress, better work-life balance, and enhanced quality of patient interaction, rather than in quantifiable increases in patient numbers or departmental income.

All of this illustrates that the productivity business case for AVT is not simple, and any decision to procure needs to carefully examine how return on investment can be achieved. There is evidence that productivity savings would be achieved, but the effects are not always directly cash-releasing, and are different for different departments and settings. A strong implementation plan would help enhance the ROI on an AVT implementation.

4. Limitations of this analaysis, opportunities for further discovery

The pilot was intentionally structured for breadth of understanding rather than experimental control, aiming to capture the diversity and complexity of clinical practice across a wide range of roles, departments, and AVT solutions. This approach has yielded a rich and nuanced picture of how AVT is experienced in real-world NHS settings, with data drawn from structured surveys, in-pilot data capture forms, and in-depth interviews. The mixed-methods design allows us to triangulate findings, combining quantitative trends with qualitative insights into workflow, user experience, and the practicalities of implementation.

The absence of randomisation, control groups, and standardised protocols means that confounding factors and selection bias have not been fully controlled for, although a range of departments and use cases were tested to provide some comparison. Much of the data is self-reported, introducing subjectivity and potential recall bias, and the heterogeneity in how AVT was used, across different clinical contexts and levels of user familiarity, makes direct comparison and aggregation challenging. As a result, while the data sets are highly effective for surfacing themes, identifying barriers and facilitators, and understanding the lived experience of clinicians, they are less robust for drawing definitive, quantitative conclusions about AVT's impact on objective outcomes such as time savings, error rates, patient safety, or clinical effectiveness.

It is also important to note that the pilot's focus was on the documentation process and staff experience, rather than direct patient outcomes or systematic error tracking. Technical limitations, such as non-integration with electronic patient records and variable IT infrastructure, further confound the assessment of AVT's potential in an optimised environment. These factors mean that the results should be seen as formative and exploratory, providing a strong foundation for further discovery and improvement, rather than as conclusive evidence of AVT's value or safety.

Looking ahead, while there are numerous opportunities for further research, the data gathered offers a comprehensive understanding of AVT's benefits and limitations in real-world clinical settings. Any implementations need to focus on careful, phased deployment and ongoing real-world evaluation to ensure that benefits are realised and any challenges are addressed. Audits and quality improvement projects should be used to optimise deployment and measure impact. This approach should allow organisations to monitor impact, refine processes, and continue learning as AVT is embedded into routine practice.

5. Recommendations

5.1 Configuration

Standardise templates where possible, but allow for personalisation

Recommendation: Any deployment should provide institutional guidance on the use of personalised versus standardised templates for letters and other documentation. Template designs and the approach to customisation should be shared during the onboarding phase for each department.

Rationale: There is a tension that needs to be resolved between two conflicting needs:

1. In general, NHS institutions suffer from a proliferation of templates and standardisation is helpful for IT management and to ensure consistency for patients.

2. AVT templates currently require a significant amount of tailoring to achieve optimal outcomes (see section 3.10 above). Error rates are higher and solutions work less effectively when AVT templates are not set up for the Large Language Models (LLMs) to respond to the specific contexts that they are deployed in.

A range of standardised templates with potential for further refinement is the ideal solution. Each department's preferred provider-designed template would be collaboratively chosen, followed by training and template iteration with individuals or groups/clinics within the department.

5.2 Deployment

Whilst the pilot does not provide unequivocal evidence for a business case, a number of lessons have been learned that would be helpful in any deployment. Recommendations below are informed directly by evaluation data, psychological acceptance models (TAM2, Trust Theory, Perceived Risk), and the observed performance of AVT tools across clinical contexts.

An open procurement would deliver on the majority requirement

Recommendation: A single standard AVT solution would serve a wide range of simple use cases across the Trust. A small number of licenses for more specialised solutions for specific more complex areas may be required to deliver on all requirements – noting however that the technology is evolving and the chosen simple solution may develop to be able to cope with more complex scenarios. A separate business case will be required to justify the cost-effectiveness of any procurement.

Rationale: This pilot showed that there were a wide range of staff groups in our Trust for whom the AVT solutions tested, in their current format, would be highly beneficial. These were largely in simpler use cases with clear templating where the main output was a patient letter or a clinical note. The products tested were broadly equivalent so an open procurement for a best-value solution would be most cost-effective. Most products on the market are working towards the level of integration that would benefit these use cases.

It is likely, although not tested here, that more complex use cases would be better served by a tailored, highly integrated solution. These more specialised areas are served by a smaller number of suppliers, therefore a small number of more costly licenses may be required to fulfil everyone's requirements. In either case, a business case will be required to justify a procurement.

Provide clear, clinician-facing information on data governance

Recommendation: Create a short, accessible explanation of how AVT data flows work (where recordings go, how long they are kept, who can access them). This directly addresses the main barrier to integrity-based trust.

Rationale: Given that trust–integrity showed only a modest increase from pre- to post-pilot, and many clinicians continued to express uncertainty about where AVT data are stored, how long they are retained, and who has access, it is clear that real-world use did not meaningfully clarify these governance questions. Providing a short, accessible explanation of AVT data flows, detailing where recordings go, how long they are kept, and the conditions under which they can be accessed, would directly address this barrier to integrity-based trust. Having an AVT product that has been officially approved through organisational governance would also help with this.

Expand deployment first in low-complexity environments, then onto complex ones

Recommendation: Begin deployment in low-complexity settings where AVT already performs reliably, using these as proof-of-concept environments. This will build early confidence among clinicians, and allow the organisation to refine templates and workflows before expanding into more demanding clinical areas. This would also be most likely to produce a clear return on investment.

Rationale: AVT performance varied markedly across clinical settings with structured, predictable environments showing high accuracy and low hallucination rates, and more complex, multi-voice environments generating more errors and requiring greater editing. A phased approach, combined with template optimisation and guidance, will reduce predictable error patterns and ensure safer performance as deployment expands.

Support timely review of outputs

Recommendation: Encourage clinicians to review AVT outputs as soon as possible after each encounter, where workload allows, so that any inaccuracies can be corrected while details remain fresh. This reinforces safe documentation practices without requiring formal changes to clinic scheduling – though may lead to a change in the way clinical and administrative time is scheduled.

Rationale: Around 40% of AVT outputs required some editing, and clinicians repeatedly noted that checking is easiest and most reliable when done immediately after the consultation. Although overall accuracy was rated highly, and the AVT tool was widely reported as being time-saving for clinicians, reviewing the output immediately after helps prevent small errors or hallucinations from entering the record and supports clinicians in using the system safely and confidently.

Maintain a focus on clinician workload and wellbeing as the core value proposition

Recommendation: Position AVT primarily as a tool that supports clinicians by reducing administrative burden, easing cognitive load, and improving overall workflow. Framing implementation around these wellbeing benefits reinforces the dimension of trust that was strongest and most stable throughout the pilot.

Rationale: Benevolence-based trust showed the largest increase post-pilot, driven by clinicians' consistent reports of reduced cognitive strain, smoother clinical encounters, and less time spent on documentation after clinic. These wellbeing gains were central to clinicians' positive evaluations of AVT and played a key role in their willingness to adopt the technology. Furthermore, interview participants frequently reflected that improved wellbeing and work life balance would very likely increase their general productivity and lead to more concrete ROI for the organisation.

Ambient Voice Technology (AVT) in Acute Services - An Evaluation

TheHill & Oxford University Hospitals NHS Foundation Trust