

GÖTEBORGS UNIVERSITET
FORSKNINGS- OCH INNOVATIONSKONTORET

Tal-till-text-funktion i socialtjänstens barnavårdsutredningar

Valideringsprojekt

Aktivitet 4: Verifieringsfråga 1

James Stålsjö

Delrapport
VFT Project#2025001
Mars, 2026

Innehållsförteckning

1. SAMMANFATTNING	1
2. BAKGRUND OCH SYFTE	2
Syfte och projektmål	2
3. GENOMFÖRANDE	4
4. UTVECKLING AV PROMPTMETODEN	6
Utveckling av AI-anpassade intervjuguidar	6
Utveckling av sammanställningsprompt	6
Sammanställningsprompt för Våldsutsatt förälder	7
Sammanställningsprompt för Våldsutövande förälder	8
Arbetsflöde för promptmetoden	8
5. RESULTAT: SAMMANSTÄLLNINGARNAS KVALITET	9
Transkriberingens betydelse för sammanställningens kvalitet	11
6. UTVÄRDERING AV PROJEKTETS NYTTA	12
7. FORTSATT UTVECKLING	13
8. SLUTSATS	14
9. REFERENSER	15
BILAGA A: AI-ANPASSADE INTERVJUGUIDER	1

1. Sammanfattning

Denna delrapport utgör den första av två rapporter i aktivitet 4 inom valideringsprojektet Tal-till-text-funktion i socialtjänstens barnavårdsutredningar. Rapporten undersöker och besvarar verifieringsfråga 1: Vilken är den rätta kvalitetsnivån för sortering och sammanställning av iRiSk-intervjuer, och vilket granskningsflöde säkerställer att kvarvarande fel är hanterbara i praktiken?

Genom iterativ prompt engineering och strukturell anpassning av iRiSk:s kortversioner av föräldraintervjuerna (Våldsutsatt respektive Våldsutövande förälder) har en ny promptmetod utvecklats för Klang AI:s Atlas-modell. Metoden består av AI-anpassade intervjuguider och två detaljerade sammanställningspromptar som tvingar modellen att strikt följa intervjuguidens avsnitts- och numreringsstruktur samt att återge 100 % ordagranna citat från transkriberingen. Inga parafraaser, sammanfattningar eller tolkningar tillåts.

Metoden testades på nio fiktiva iRiSk-intervjuer. Resultaten visar en markant förbättrad kvalitetsnivå jämfört med flödesmetoden i aktivitet 3: sammanställningarna behåller exakt struktur och innehåller enbart verifierbara citat. Samtidigt kvarstår stokastisk variation i outputen, vilket kräver flera körningar och manuell granskning. Transkriberingens kvalitet visade sig vara avgörande för slutresultatet; bristfällig talaridentifikation och oupptäckta transkriptionsfel ökar felmängden.

Verifieringsfrågan är besvarad, men de definierade utfallskriterierna för positivt resultat uppnås endast delvis. Den höjda kvalitetsnivån kräver omfattande manuell bearbetning av transkriberingar, vilket motverkar den tidsbesparing som var projektets ursprungliga syfte. Rapporten ger därmed ett underlag för beslut om eventuell småskalig skarp pilottest i ordinarie verksamhet, samtidigt som fortsatt teknisk utveckling rekommenderas för att nå stabil och tidsbesparande användbarhet.

2. Bakgrund och syfte

Detta är den första av två rapporter i den fjärde fasen av valideringsprojektet Tal-till-textfunktion inom socialtjänstens barnavårdsutredningar vid Forsknings- och Innovationskontoret, Göteborgs universitet. Valideringsprojektets övergripande mål är att utveckla ett digitalt tal-till-textstöd med artificiell intelligens (AI) som automatiskt transkriberar intervjusamtal till strukturerad text, för användning vid strukturerade risk- och skyddsbedömningar med metodstödet iRiSk (Eriksson m.fl., 2025) inom socialtjänsten. Idén är att AI-verktyget ska ersätta manuella anteckningar i utredningssamtal och därmed minska risken för att väsentlig information går förlorad, förbättra bedömningsunderlaget och utredningarnas kvalitet, samt frigöra tid som ökar socialsekreterarens närvaro i klientsamtalen. Denna rapport fokuserar på kvalitetsnivån i AI-producerade sammanställningar av fiktiva iRiSk-intervjuer.

I projektets föregående fas (aktivitet 3) undersöktes användbarheten av ett sådant tal-till-textverktyg från tjänsteleverantören Klang AI. Under en sex månader lång testperiod fick 16 socialsekreterare från sju kommuner testa Klangs AI-verktyg i samband med fiktiva iRiSk-intervjuer (Stålsjö & Lennartsson, 2026). Resultaten visade att socialsekreterarna tror att AI kan höja deras arbetseffektivitet och är användbart vid barnavårdsutredningar, framför allt genom tidsbesparingen som den automatiserade dokumentationen medför. Testet identifierade därtill flertalet brister med Klangs AI-verktyg som behöver åtgärdas innan en skarp testpilot av iRiSk med AI-stöd inom socialtjänstens ordinarie verksamhet är aktuell.

Ett avgörande problem är att generativa AI-modeller som Klangs Atlas-modell riskerar att producera missvisande resultat (hallucinera) om modellen tillåts tolka intervjumaterial utan tydliga restriktioner, något som Integritetsskyddsmyndigheten (IMY) har varnat för (2025). På grund av detta har vi beslutat att inte använda Klangs inbyggda flödesfunktion Flow för sammanställning av iRiSk-intervjuer i nuläget. Krav på nära 100% ordagranna citat från transkriberingen är i dagsläget en förutsättning för att sammanställningen ska kunna användas som tillförlitligt beslutsunderlag i barnavårdsutredningar. Detta krav har ställts av forskargruppen bakom iRiSk. Eftersom Klangs Flow för iRiSk inte når upp till dessa krav så finns det ett behov av att undersöka kapaciteten i Klangs sammanställningsfunktion Atlas närmare.

Syfte och projektmål

Uppdraget i aktivitet 4 syftar till att skapa förutsättningar för en småskalig testpilot i socialtjänstens ordinarie barnavårdsutredningar.

Delrapport 1 undersöker och besvarar verifieringsfråga 1: *Vilken är den rätta kvalitetsnivån för sortering och sammanställning av iRiSk-intervjuer och vilket granskningsflöde säkerställer att kvarvarande fel är hanterbara i praktiken?*

Målet är att höja kvaliteten i Klang AI:s sammanställningar av iRiSk-intervjuer genom modifieringar av filtrering och prompting eller ett bättre arbetsflöde och tydligare arbetsrutiner. I samråd med forskargruppen bakom iRiSk definierades den rätta kvalitetsnivån som att sammanställningen ska återge intervjuguidens exakta struktur i sin helhet – inklusive alla avsnitt och numrerade frågor – samt att varje fråga besvaras med ordagranna citat från transkriberingen. Inga parafrafer, sammanfattningar eller AI-genererade tolkningar får förekomma. Ett positivt utfall innebär att denna nivå uppnås samtidigt som transkriberingskvaliteten för intervjuerna förblir stabil.

Detta innebär att en ny promptmetod för sammanställning av intervjumaterial med Klang AI behöver utvecklas, vilket innebär att:

- Omarbeta de befintliga föräldraintervjuerna med iRiSk (Våldsutsatt/Våldsutövande) till AI-anpassade versioner som underlättar filavläsningen för Klang AI att korrekt tolka och replikera intervjuguidens struktur.
- Utveckla ett deterministiskt konfigurerat sammanställningsprompt med exakta instruktioner om hur intervjumaterialet ska bearbetas och struktureras.
- Testa den utvecklade promptmetoden med fiktiva iRiSk-intervjuer och utvärdera kvaliteten i sammanställningarna som metoden producerar.

Om de strukturerade resultaten från iRiSk-intervjuerna förblir opålitliga eller utgör en oacceptabel risk för dokumentationens kvalitet, även efter mänsklig granskning, kan ett andra alternativ vara att utveckla ett dedikerat AI-verktyg med hjälp av IT-utvecklare eller entreprenör.

3. Genomförande

Arbetet bestod av en iterativ metodutveckling där prompt engineering kombinerades med strukturell anpassning av de befintliga kortversionerna av intervjuguiderna för Våldsutsatt förälder och Våldsutövande förälder (Broberg m.fl., 2026a, 2026b). Syftet med dessa anpassningar var att testa och förbättra sammanställningsfunktionen i Klang AI så att den överensstämde med iRiSk-strukturen och uppfyllde kravet på ordagranna citat.

Nio rollspelade iRiSk-intervjuer analyserades. Dessa intervjuer hade spelats in i förväg och genomförts av handläggare vid Familjerätten Hägersten-Älvsjö i Stockholm. Under intervjuerna användes kortversionerna av föräldrantervjuguiderna. Dessa kortversioner hade anpassats särskilt för användning tillsammans med tal-till-textstöd och utan krav på manuella anteckningar. Intervjuguiderna testades efter att personal från Familjerätten besökte forskargruppen bakom iRiSk vid Psykologiska institutionen, Göteborgs universitet, i december 2025. Ljudfilerna tillhandahölls av Ole Hultmann.

Ljudfilerna bearbetades i Klang AI med funktionerna i modellerna Capture och Atlas (för en närmare beskrivning av dessa modeller, se Stålsjö & Lennartsson, 2026). Åtta av intervjutranskriptionerna genomgick en översiktlig bearbetning med Klangs inbyggda redigeringsfunktion för att automatiskt identifiera transkriptionsfel. Dessa fel korrigerades manuellt i redigeringsläget, vilket tog cirka 2–3 minuter per transkribering, för att minimera felkällor i de efterföljande sammanställningarna. För att ytterligare öka resultatens tillförlitlighet och minska variationen i sammanställningarna till följd av transkriptionsfel valdes den nionde intervjun (Våldsutsatt förälder) ut som referensintervju på förhand. Referensintervjun valdes ut eftersom den var väl genomförd, hade en hanterbar längd på 24 minuter och 15 sekunder vilket underlättade en noggrann manuell genomgång, samt innehöll både en manlig och en kvinnlig röst – en kombination som enligt erfarenheter från aktivitet 3 ger betydligt färre transkriptionsfel. Transkriberingen av referensintervjun validerades manuellt mot ljudfilen för att säkerställa fullständig korrekthet. Bearbetningen av denna intervju tog cirka 45 minuter, vilket motsvarar en genomsnittlig tidsåtgång på 1 minut och 51 sekunder per inspelad minut. Referensintervjun användes därefter som utgångspunkt för metodutvecklingen, både för förbättring och validering av prompter och intervjuguide, och som referenspunkt när kvaliteten i sammanställningarna av de övriga åtta intervjuerna bedömdes.

Anpassningen av intervjuguiderna för AI-användning var ursprungligen inte planerad, men visade sig vara en avgörande förutsättning för att Atlas-modellen skulle kunna producera tillförlitliga sammanställningar. De ursprungliga Word-filerna innehöll en stor mängd störande information, såsom instruktioner, kommentarer och annan vägledning riktad till handläggaren. Denna information skapade brus som hindrade modellen från att korrekt identifiera och följa intervjuguidens struktur. Genom att rensa och standardisera

dokumenterna – bland annat genom enhetlig formatering, konsekvent numrering och borttagning av all icke-relevant text – uppnåddes betydligt bättre resultat. Denna anpassningsprocess krävde endast tre iterationer och var både okomplicerad och tidsmässigt effektiv. Utvecklingen av sammanställningsprompterna krävde 16 iterationer med upprepade tester av varje version för att nå önskad precision.

Metodutvecklingen följde ett konsekvent och repetitivt arbetsflöde i Klang AI. Detta innebar att den AI-anpassade intervjuguiden först laddades upp, därefter kördes sammanställningsprompten, varefter outputen granskades och validerades manuellt. Slutligen anpassades antingen intervjuguiden eller prompten utifrån granskningen. Arbetet genomfördes under februari och mars 2026 med stöd av xAI:s Grok 4.20 (2026) för promptoptimering och AI-anpassning av intervjuguiderna.

4. Utveckling av promptmetoden

I detta avsnitt presenteras resultatet av promptmetodens utveckling: de AI-anpassade intervjuguiderna, de slutgiltiga sammanställningspromptarna samt det praktiska arbetsflödet för användning i Klang AI.

Utveckling av AI-anpassade intervjuguides

Anpassningen av iRiSk:s korta föräldrintervjuer för Våldsutsatt respektive Våldsutövande förälder syftade till att skapa AI-vänliga dokument som underlättar maskinell avläsning och ökar Atlas-modellens förmåga att korrekt identifiera, följa och reproducera intervjuguidens exakta struktur och numrering i sammanställningarna. De fullständiga versionerna av dessa strukturellt anpassade intervjuguides lämnas inte ut i denna delrapport på grund av upphovsrättsliga skäl. De kan på begäran tillhandahållas forskare och kliniker som aktivt deltar i vidareutvecklingen av metodstödet iRiSk (se Bilaga A).

De AI-anpassade versionerna blev betydligt mer koncisa än de ursprungliga filerna. Guiden för våldsutsatt förälder reducerades från 2014 till 831 ord, vilket motsvarar en minskning med cirka 59 % (1 183 ord). Antalet numrerade frågor ökade samtidigt från 61 till 65 genom att skriva ut villkorade frågor. Guiden för våldsutövande förälder kortades från 3 014 till 1 012 ord, en reduktion med cirka 66 % (2 002 ord), medan antalet frågor förblev oförändrat på 69.

En central anpassning bestod i att tydligt skilja avsnitt och intervjufrågor åt. Avsnitten (A–M respektive A–F) presenterades med ett enhetligt och dedikerat rubriktypsnitt och bibehållen bokstavsnumrering, medan alla intervjufrågor angavs i samma typsnitt och textstorlek genom hela dokumentet, med fetstilt siffrumrering och bevarad ordningsföljd. Radavståndet sattes till 1,15, indrag anpassades och visuell variation minimerades för att optimera maskinell läsbarhet. Dessa formaterings- och strukturella justeringar, tillsammans med mer entydiga och kortfattade frågeformuleringar, ökade Atlas-modellens förmåga att korrekt återskapa intervjuguidens struktur i sammanställningarna jämfört med de obearbetade kortversionerna.

Utveckling av sammanställningsprompt

Två detaljerade sammanställningsprompter togs fram för att instruera Atlas-modellen att enbart använda exakta citat från transkriberingen och strikt replikera intervjuguidens struktur och numrering. De slutgiltiga promptarna för intervjuguiderna Våldsutsatt förälder och Våldsutövande förälder presenteras nedan.

Sammanställningsprompt för Våldsutsatt förälder

Du ska sammanställa @transkriberingen strikt enligt exakt samma avsnitt (A-M) och numrering (1-65) som i @intervjuguiden.

Gå igenom varje avsnitt (A-M) och varje numrerad fråga (1-65) i den ordning de står i @intervjuguiden. Avsnitt A-M med frågenummer: A. Relationen till barnet (1-8), B. Förälderns beskrivning av våld föräldern/barnet utsatts för (9-11), C. Risk/skyddsaspekter - barn (12-14), D. Våld och andra allvarliga händelser som barnet utsatts för (15-22), E. Ert barns boende (23-30), F. Säkerhetsplanering (31-36), G. Risk-skyddsaspekter - förälder (37-41), H. Juridik (42-47), I. Psykisk ohälsa och missbruk (48-50), J. Positiv kontakt med barnet (51), K. Talar man om våldet med barnet (52-57), L. Förälderns uppfostringsstrategier (58-63), M. Avslutning (64-65). Varje fråga måste skrivas ut.

För varje fråga:

- Skriv först **Handläggare:** följt av den exakta formulering som handläggaren faktiskt använde i transkriptionen (100 % ordagrant, även om den är lite annorlunda från guiden).
- Direkt under: **Förälder:** följt av förälderns svar som 100 % ordagrant citat från @transkriberingen (ingen sammanfattning, ingen parafras. Kopiera identiskt).
- Eventuella följdfrågor återges därefter i samma Handläggare:/Förälder:-format direkt under huvudfrågans svar utan egna nummer eller rubriker.

Om en fråga aldrig ställdes alls: skriv **Frågan ej ställd**

Om en fråga aldrig besvarades: skriv **Frågan ej besvarad**

Behåll alla avsnitt A-M som rubriker. Exakt som i @intervjuguiden. Inga extra kommentarer utanför strukturen.

Skapa ett Word-dokument.

Sammanställningsprompt för Våldsutövande förälder

Du ska sammanställa @transkriberingen strikt enligt exakt samma avsnitt (A-F) och numrering (1-69) som i @intervjuguiden.

Gå igenom varje avsnitt (A-F) och varje numrerad fråga (1-69) i den ordning de står i @intervjuguiden. Avsnitt A-F med frågenummer: A. Relationen till barnet (1-11), B. Förälderns mående och beteende (12-20), C. Boende, vårdnad, umgänge/kontakt (21-37), D. Våld (38-63), E. Förälderns behov av stöd (64-66), F. Avrundning (67-69). Varje fråga måste skrivas ut.

För varje fråga:

- Skriv först ****Handläggare:**** följt av den exakta formulering som handläggaren faktiskt använde i transkriptionen (100 % ordagrant, även om den är lite annorlunda från guiden).
- Direkt under: ****Förälder:**** följt av förälderns svar som 100 % ordagrant citat från @transkriberingen (ingen sammanfattning, ingen parafraas. Kopiera identiskt).
- Eventuella följdfrågor återges därefter i samma Handläggare:/Förälder:-format direkt under huvudfrågans svar utan egna nummer eller rubriker.

Om en fråga aldrig ställdes alls: skriv ****Frågan ej ställd****

Om en fråga aldrig besvarades: skriv ****Frågan ej besvarad****

Behåll alla avsnitt A-F som rubriker. Exakt som i @intervjuguiden. Inga extra kommentarer utanför strukturen.

Skapa ett Word-dokument.

Arbetsflöde för promptmetoden

För att använda promptmetoden krävs att ljudfilerna först har laddats upp till Klang AI. Därefter navigerar man till arbetsytan för den aktuella intervjun och följer ett strukturerat arbetsflöde. Transkriberingsfilen bearbetas manuellt så noggrant som möjligt och namnen på intervjupersonerna döps om till "Handläggare" och "Förälder". Den AI-anpassade versionen av intervjuguiden infogas i arbetsytan i Word- eller .docx-format. Båda filerna döps därefter om till "transkriberingen" respektive "intervjuguiden" för att matcha promptens referenser. Prompten infogas i dialogrutan och filreferenserna kontrolleras innan prompten körs. Den genererade sammanställningen granskas sedan med avseende på struktur och innehåll. Om resultatet är acceptabelt kan Word-dokumentet sparas direkt, eventuella fel korrigeras manuellt eller processen upprepas för att erhålla en bättre version.

Vid vissa körningar misslyckas Klang med att producera sammanställningen som ett nedladdningsbart Word-dokument. I dessa fall underlättar det att ta bort instruktionen "Skapa ett Word-dokument" från prompten och i stället använda Klangs inbyggda funktion för att skapa dokument via drop-down-menyn i dialogrutan. Det är viktigt att endast @transkriberingen, @intervjuguiden och eventuellt dokumentfunktionen läggs till som sammanhang inför varje ny körning.

5. Resultat: Sammanställningarnas kvalitet

Med den nya promptmetoden som kombinerar AI-anpassade intervjuguider och dedikerade promptar är det möjligt att producera sammanställningar av iRiSk-föräldrantervjuer med hög strukturell träffsäkerhet som enbart använder citat direkt från transkriberingarna. Atlas-modellen genererade dokument med konsekvent formatering, tydliga avsnittsrubriker och numrerade intervjufrågor. Svaren på varje fråga presenterades i det föreskrivna formatet – *Handläggare*: följt av den exakta frågeformuleringen och *Förälder*: följt av det exakta svaret – och uteblivna frågor eller svar noterades tydligt.

Atlas-modellen, likt alla stora språkmodeller, är dock stokastisk (slumpmässig) (Su, 2025). Därför uppnåddes den höga kvalitetsnivån sällan vid första körningen trots identiska indata (transkribering, intervjuguide och prompt). Promptmetoden visade en markant variationsbredd i outputen och krävde ofta flera körningar för att producera en acceptabel sammanställning. Variationen i kvalitet var dock kraftigt begränsad jämfört med den tidigare flödesmetoden.

För att illustrera både förbättringen gentemot flödesmetoden från aktivitet 3 och den variation som fortfarande förekommer presenteras ett representativt exempel från referensintervjun i Box 1 på nästa sida. Exemplet avser fråga 65 i avsnitt M Avslutning från intervjuguiden för Våldsutsatt förälder (nummer 63 i den version som användes i aktivitet 3).

Som framgår av exemplet eliminerade den nya promptmetoden helt parafrasering och AI-genererad tolkning. Alla citat var identiska med transkriberingen i samtliga nio intervjuer. De två prompt-körningarna visar samtidigt den kvarvarande slumpmässiga variationen: körning 1 resulterade i felaktig frågenummering men behöll mer kontext, medan körning 2 följde intervjuguidens struktur exakt men gav snävare uttalanden. Felaktig nummering kunde åtgärdas enkelt vid manuell granskning. Variationen i kontext/följdfrågor riskerar däremot att utelämna väsentlig information och kräver större ansträngning för att åtgärdas.

Ett annat exempel från referensintervjun uppstod i avsnitt L om uppfostringsstrategier. Handläggaren inledde avsnittet genom att säga *”Vi har inte fyllt i [skattningsformuläret] om uppfostringsstrategier, så vi hoppar över det [...]”*. Dock följdes påståendet utan fördröjning upp med frågan *”[...] Men skulle du säga att du och Micke kan samarbeta kring omsorg och uppfostran av Bruno?”*. Denna fråga tillhör avsnitt L, men ignorerades i de flesta körningar av Klang, som tolkade uttalandets första del som instruktioner att direkt gå vidare till nästa avsnitt. Endast i några få körningar registrerades den korrekta sekvensen i sammanställningen, med intakt struktur och fullständiga citat från transkriberingen.

Ytterligare, men mindre vanliga variationer över intervjuerna, var att följdfrågor ibland presenterades som numrerade frågor, att uttalanden slumpmässigt markerades med

Box 1 Jämförelse av output mellan sammanställningar från referensintervjun.

Exempel: Fråga 65 – Vad gör barnet och du tillsammans som barnet tycker om?

Transkribering (rådata)

Handläggare: Och vi pratade om det här tidigare, men jag tänkte ställa frågan igen. Vad gör du och barnet tillsammans som barnet tycker om?

Förälder: Ja, men idag ska vi väl gå och fika tänkte jag. Och det tycker han ju också är lite kul att få en kanelbulle och sådär.

Handläggare: Ja, men det låter som att ni kommer ha en fin dag idag, du och Bruno. Tack för att du kom.

Flöde (Aktivitet 3)

63. Gör barnet och du något tillsammans som barnet tycker om? (Alternativ: Nej, Ja)

Ja

Om ja, beskriv

Bruno uppskattar vardagshändelser som att gå till affären och handla, leka i lekparken och sandlådan. Han tycker om när vi träffar min syster med hennes barn och när vi åker till mormor och sover där. Idag planerar vi att gå och fika.

Gör ni något mer? (Alternativ: Enbart aktiviteter, Vardaglig gemenskap)

Vardaglig gemenskap

Hur ofta gör ni något tillsammans som barnet uppskattar?

Ofta, nästan jämt umgås vi.

Prompt – körning 1

62. Vad gör barnet och du tillsammans som barnet tycker om?

Handläggare: Och vi pratade om det här tidigare, men jag tänkte ställa frågan igen. Vad gör du och barnet tillsammans som barnet tycker om?

Förälder: Ja, men idag ska vi väl gå och fika tänkte jag. Och det tycker han ju också är lite kul att få en kanelbulle och sådär.

Handläggare: Ja, men det låter som att ni kommer ha en fin dag idag, du och Bruno. Tack för att du kom.

Prompt – körning 2

65. Vad gör barnet och du tillsammans som barnet tycker om?

Handläggare: Och vi pratade om det här tidigare, men jag tänkte ställa frågan igen. Vad gör du och barnet tillsammans som barnet tycker om?

Förälder: Ja, men idag ska vi väl gå och fika tänkte jag. Och det tycker han ju också är lite kul att få en kanelbulle och sådär.

citationstecken och att direktlänkar till transkriberingen inte fungerade med felmeddelandet "Vald tid överstiger transkriberingens längd". Ett återkommande tekniskt problem var att Klang ofta avbröt genereringen halvvägs, vilket medförde onödigt extra arbete.

Transkriberingens betydelse för sammanställningens kvalitet

När de åtta övriga intervjuerna analyserades framträdde tydliga effekter av bristfälliga transkriberingar. Dessa transkriberingar hade endast bearbetats med Klangs automatiska redigering och sammanställningarna innehöll därför fler fel än referensintervjun.

Klang hade påtagliga svårigheter att automatiskt identifiera potentiella transkriberingsfel med den inbyggda redigeringsfunktionen. I de flesta intervjuer flaggades färre än 10 uttalanden för granskning, men det verkliga antalet var i själva verket mycket högre. Ett sådant exempel var att frågan "*Har barnen någon kontaktfamilj?*" i en av intervjuerna obemärkt transkriberades som "*Har barnen någon kontakt från JIP?*", trots att de två fraserna är fonetiskt mycket olika. I vissa fall hade Klang även svårt att identifiera uttalanden som tydligt hördes i ljudfilen men uteblev från transkriberingen till synes utan någon anledning. Återkommande från aktivitet 3 var ett märkbart problem med felaktig talaridentifikation vid två kvinnoröster.

Transkriberingsfelen var inte begränsade till dåliga ljudupptagningsförhållanden, utan ökade i frekvens vid överlappande tal, språkbrytning och oljud från bläddrande papper.

I sammanställningarna medförde dessa fel att variationen i struktur och fullständighet ökade markant. Fler uttalande tillskrevs fel person vilket resulterade i att fler frågor felaktigt angavs som ej ställda eller redovisades med missvisande svar, även om citatnoggrannheten i sig förblev intakt. Transkriberingarnas träffsäkerhet är därmed av större betydelse för sammanställningarnas kvalitet vid användning av promptmetoden, jämfört med flödesmetoden.

6. Utvärdering av projektets nytta

Verifieringsfråga 1 ställde frågan om vilken den rätta kvalitetsnivån är och vilket granskningsflöde som gör kvarvarande fel säkra i praktiken. Rätt kvalitetsnivå specificerades i samråd med forskargruppen bakom iRiSk (se avsnitt 2). Ett positivt utfall definierades som att sammanställningarna återger exakta citat, följer intervjuguidens struktur och att transkriberingskvaliteten förblir stabil.

Med promptmetoden kan Atlas-modellen producera sammanställningar som stämmer överens med strukturen i intervjuguiden och enbart använder citat från transkriberingen för att besvara intervjufrågorna. Till skillnad från flödesmetoden i aktivitet 3, där Atlas kunde "gissa sig till" essensen i konversationerna genom statistisk slutledning (Bender m.fl., 2021), belägger den nya promptmetoden modellen med strikta deterministiska konfigurationer. Den kvarvarande slumpmässiga variationen hanteras genom ett granskningsflöde: (1) noggrann manuell bearbetning av transkriberingen i Klang, (2) kontroll av filreferenser i prompten samt (3) rutinmässig granskning av den genererade sammanställningen.

I praktiken är detta hursomhelst en ansträngande procedur som framför allt kräver att användaren säkerställer att transkriberingen träffsäkert återger intervjusituationen. Klangs inbyggda redigeringsfunktion är inte tillräcklig för att automatisk identifiera potentiella talfel och kräver omfattande manuell bearbetning som motverkar valideringsprojektets syfte att vara tidsbesparande. Emedan promptmetoden höjer sammanställningarnas kvalitet jämfört med flödesmetoden, så kan inte transkriberingarnas kvalitet sägas vara stabil.

Verifieringsfrågan är därmed besvarad men kriterierna för positivt utfall uppnås endast delvis.

7. Fortsatt utveckling

Förutsättningen för att sammanställa frågor och svar från iRiSk-intervjuer med citat från handläggare och förälder är träffsäkra transkriberingar.

Högre kvalitet på transkriberingarna kan uppnås på flera sätt. Handläggarna som utför iRiSk-intervjuerna bör få tydliga instruktioner om att inte tala i mun på intervjupersonen. Frågorna i intervjuguiden ska ställas klart och tydligt. Man bör uppmuntra växelvis turtagning under samtalet. Oljud som stör ljudinspelningen ska minimeras, till exempel bläddrande bland papper. Slutligen bör ändamålsenlig inspelningsutrustning användas, exempelvis bättre mikrofoner, för att förbättra ljudupptagningen. Dessa åtgärder minimerar antalet fel i transkriberingen, underlättar den manuella bearbetningen och är i slutändan tidsbesparande. Att manuellt bearbeta transkriberingarna är hur som helst fortfarande en tidskrävande process. Den skulle underlättas ytterligare genom tekniska förbättringar.

Vid just bearbetning av iRiSk-intervjuer skulle det underlätta om Klang AI kunde kalibrera sin inbyggda redigeringsfunktion. Funktionen bör bli känsligare vid automatisk identifikation av potentiella transkriberingsfel. Det bör även undersökas om de automatiska transkriberingarna kan göras mer träffsäkra. Detta kan ske om Klang tillhandahåller ljudinspelningar och högkvalitativa transkriberingar från iRiSk-intervjuer. Då kan modellen finjusteras för att bli mer domänspecifik (Radford m.fl., 2022).

Vad beträffar kvaliteten på outputen (sammanställningarna) kan variationen i resultaten möjligtvis minimeras ytterligare om prompten för sammanställning och intervjuguiderna struktureras enligt andra format eller mallar, exempelvis JSON (He m.fl., 2024). Prestandan och outputkvaliteten från stora språkmodeller är också beroende av uppgiftens omfattning och antalet tokens som AI-modellen kan hantera (Liu m.fl., 2023; Du m.fl., 2025). Inför fortsatt utvecklingsarbete är det nödvändigt att ta reda på AI-modellens begränsningar för att kunna avgöra om modellen kan hantera de ålagda uppgifterna effektivt (Paulsen, 2025). Dessa insikter kan motarbeta behovet av flera körningar för att erhålla godtagbara resultat.

Långsiktigt behöver forskargruppen producera en fullständig iRiSk-utredning som lever upp till "gold standard" och representerar vad de vill uppnå med utvecklingen av en AI-tjänst som är fullkomligt integrerad med iRiSk. Detta är av yttersta vikt. En sådan exempelutredning informerar om vilka krav en ändamålsenlig AI-modell behöver uppfylla och vägleder valen i det fortsatta utvecklingsarbetet.

8. Slutsats

Genom att använda en kombination av AI-anpassade intervjuguider och dedikerade sammanställningspromptar är det möjligt att producera sammanställningar av iRiSk-föräldrainsatser som strikt följer intervjuguidens struktur och enbart innehåller ordagranna citat från transkriberingen. Den nya promptmetoden ger därmed en markant högre kvalitetsnivå jämfört med flödesmetoden som testades i aktivitet 3. Verifieringsfråga 1 är därmed besvarad.

Samtidigt uppnås de definierade utfallskriterierna för ett positivt resultat endast delvis. Den ökade kvaliteten kräver betydligt mer manuell bearbetning av transkriberingarna och flera körningar av prompten, vilket äter upp den tidsbesparingen och den förbättrade användbarhet som rapporterades i föregående fas (Stålsjö & Lennartsson, 2026).

Avslutningsvis finns förutsättningar för att ta projektet vidare till ett småskaligt skarpt pilottest i socialtjänstens ordinarie verksamhet med vuxna/föräldrainsatserna. Med den utvecklade promptmetoden kan vi förvänta oss högre kvalitet på sammanställningarna jämfört med flödesmetoden, men på bekostnad av den förväntade tidsbesparingen. Användningen av AI-verktyget kommer med nuvarande tekniska förutsättningar sannolikt att upplevas som trög. Huruvida ett pilottest är aktuellt under dessa villkor är avhängigt av syftet med testet, som ännu inte har specificerats. Det behöver därför beslutas om ett pilottest ska påbörjas i nästa steg eller om tekniken först måste förbättras ytterligare. Därtill bör syftet med ett sådant pilottest formuleras tydligt: ska det primärt samla in data som underlag för träning av en mer domänspecifik AI-modell, eller endast observera AI:ns kapacitet i skarpa sammanhang och fånga handläggarnas erfarenheter på nytt?

9. Referenser

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Broberg, A., Hultmann, O., Axberg, U., & Eriksson, M. (2026a). *iRisk: Risk- och skyddsintervju: Våldsutsatt förälder (3e omarbetade upplagan, Kortversion)*. Göteborgs universitet.
- Broberg, A., Hultmann, O., Axberg, U., Eriksson, M., Möller, C., & Andersson Sjøe, M. (2026b). *iRisk: Risk- och skyddsintervju: Våldsutövande förälder (3e omarbetade upplagan, Kortversion)*. Göteborgs universitet.
- Du, Y., Tian, M., Ronanki, S., Rongali, S., Bodapati, S., Galstyan, A., Wells, A., Schwartz, R., Huerta, E. A., & Peng, H. (2025). Context length alone hurts LLM performance despite perfect retrieval. arXiv. <https://doi.org/10.48550/arXiv.2510.05381> [Preprint].
- Eriksson, M., Axberg, U., Broberg, A., Chawinga, E., Hultmann, O., & Iversen, C. (2025). *iRisk insatser & risk-skyddsbedömningar för våldsutsatta barn: Användarstöd iRisk-intervjuer (3:e omarbetade upplagan)*. Psykologiska institutionen, Göteborgs universitet.
- He, J., Rungta, M., Koleczek, D., Sekhon, A., Wang, F. X., & Hasan, S. (2024). *Does prompt formatting have any impact on LLM performance?* arXiv. <https://doi.org/10.48550/arXiv.2411.10541> [Preprint].
- Integritetsskyddsmyndigheten. (2025, 8 september). *Yttrande i förhandssamråd enligt dataskyddsförordningen avseende behandling av personuppgifter i AI-tjänst inom socialnämnden och arbetsmarknadsnämnden i Helsingborgs kommun* (Diariennr IMY-2025-12866) [Yttrande].
- Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2023). Lost in the middle: How language models use long contexts. arXiv. <https://doi.org/10.48550/arXiv.2307.03172> [Preprint].
- Paulsen, N. (2025). Context is what you need: The maximum effective context window for real world limits of LLMs [Preprint].
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. arXiv preprint arXiv:2212.04356. <https://arxiv.org/pdf/2212.04356.pdf>

- Stålsjö, J., & Lennartsson, E. (2026). *Tal-till-text-funktion i socialtjänstens barnavårdsutredningar: Innovationsprojekt iRiSk med Klang AI* (Slutrapport Project#2025001). Göteborgs universitet, Forsknings- och innovationskontoret.
- Su, W. (2025). Do large language models (really) need statistical foundations? arXiv. <https://doi.org/10.48550/arXiv.2505.19145> (Accepted at The Annals of Applied Statistics)
- xAI. (2026). Grok 4.20 (Version mars 2026) [Stor språkmodell]. <https://x.ai/grok>

Bilaga A: AI-anpassade intervjuguider

De fullständiga versionerna av dessa strukturellt anpassade intervjuguider lämnas inte ut i denna delrapport på grund av upphovsrättsliga skäl. Forskare och kliniker som aktivt deltar i vidareutvecklingen av metodstödet iRisk kan begära tillgång till dokumenten genom att kontakta Ole Hultmann eller Anders Broberg vid Psykologiska institutionen, Göteborgs universitet.

Kontaktuppgifter:

Ole.hultmann@psy.gu.se

Anders.broberg@psy.gu.se