

Rapport

HyresAI

Utvärdering av två metoder för att extrahera uppgifter från ostrukturerade dokument



EKONOMISTYRNINGSVERKET

Publikationen kan laddas ner
från ESV:s webbplats esv.se.

Datum: 2024-03-04

Dnr: 2024-00649-1

ESV-nr: 2024:24

Copyright: ESV

Rapportansvarig: Pierre Mesure

Förord

Denna rapport beskriver ett internt projekt inom Ekonomistyrningsverket (ESV) i vilket ESV:s datalabb har utforskat om stora språkmodeller och andra typer av modeller kan vara ett stöd för att extrahera uppgifter från hyresavtal.

Projektgruppen har bestått av utredarna Hannes Jacobsson och Pierre Measure.

Bozena Lilja och Björn Andersson på enheten Analys och prognos har bidragit med viktig kunskap i projektet.

Stockholm

2024-03-04

Martin Sparr

Enhetschef

Pierre Measure

Utredare

Innehåll

Sammanfattning	5
1 Beskrivning av datamängden	6
2 Beskrivning av modellerna	9
2.1 Donut	9
2.2 GPT-4-Vision	10
3 Metodbeskrivning	11
3.1 Framtagning av en test-datamängd	12
3.2 Donuts finjustering	12
3.3 Instruktionsoptimering för GPT-4	13
4 Resultat	14
4.1 Donut	15
4.1 GPT-4-Vision	16
5 En snabb teknisk utveckling	18
6 Slutsatser	20

Sammanfattning

Under hösten 2023 har vi i Ekonomistyrningsverkets (ESV:s) datalabb utfört tester med två modeller för att extrahera information från ostrukturerade dokument. Detta är en vanligt förekommande uppgift i myndigheten och kräver än så länge manuellt arbete. I datalabbet har vi många gånger använt oss av stora språkmodeller för att analysera och extrahera information ur ren och kurerad löptext. Det nya med den uppgift vi här testat är att modellerna också behöver använda visuell information för att kunna tolka och extrahera rätt uppgifter.

I vårt test har vi använt oss av de hyresavtalsdokument som ligger till grund för myndighetens databas över hyreskontrakt, KAR. Underlagen kommer som pdf-bilagor per mejl och innehåller uppgifter över exempelvis hyresvärd, adress, lokalarea och hyra. Det mesta av informationen som avkodas ligger på en sida, utformad som ett formulär. För en människa är kontrakten lätta att läsa och informationen enkel att transkribera. Men för en maskin är det svårare.

Denna rapport visar ändå att ny teknik redan kan uppnå en ganska hög nivå av pålitlighet, och skulle kunna användas som stöd för att effektivisera transkriberingsarbetet. Vi beskriver mer i detalj vilka svårigheter de testade modellerna möter och hur en bättre databeredskap skulle kunna hjälpa dem.

1 Beskrivning av datamängden

ESV ansvarar för att samla in och sammanställa uppgifter om den statliga lokalförsörjningen. De statliga myndigheterna är skyldiga att skicka in uppgifter om ingångna hyresavtal samt förändringar i dessa till ESV. Varje år mottar ESV cirka 1 500 kontrakt, varav 500 är nya och resterande avser förändringar eller avslutade kontrakt. Uppgifterna registreras löpande i databasen Kontrakt- och användarregister (KAR).

Myndigheterna skickar in uppgifterna per mejl, med kontrakten bifogade som pdf-filer. Dessa sparas manuellt ned på disk, och uppgifterna extraheras manuellt direkt in i KAR. KAR är en intern webbapplikation som består av ett sökgränssnitt för att hitta ett avtal och en avtalssida där man kan ta del av ett avtals uppgifter och redigera dem.

Figur 1: Avtalssidan i webbapplikationen KAR

Kontrakt:
 AB29960004 [PDF](#)
Hyresvärd:
 Fabege AB ...
Fastighet:
 Barnhusväderkvarnen 36 ...
Adress:
 Drottninggatan 89
Kommun:
 Stockholm
Tecknare:
 Ekonomistyrningsverket ...
Tecknarens referens:
 Ht 1
Hyresvärdens referens:
 5410-1012,4
 Bevakat avtal för utlämning:

Kontrakt nr	Tecknare	Datum
AB29960001	Ekonomistyrningsverket	2020-02-01
AB29960002	Ekonomistyrningsverket	2020-02-01
AB29960003	Ekonomistyrningsverket	2020-02-01
✓ AB29960004	Ekonomistyrningsverket	2020-03-01
AB29960005	Ekonomistyrningsverket	2020-03-01
AB29960006	Ekonomistyrningsverket	2020-03-01

Giltighet

Avtalstid fr.o.m. 2020-03-01 **T.o.m.** 2025-02-28 **Uppsagt**
Förlängning 36 Mån **Uppsägning** 12 Mån **Slutdatum** 2025-02-28

Areauppgifter

Ovan mark	2 718	Butik	0	Arkiv	0
Under mark	0	Kontor	2 718	Undervisning	0
Totalt	2 718	Förråd	0	Övrigt	0

Registrera

Komponenter
 Brukare

När en myndighet skickar in nya filer som avser uppdateringar eller avslut av kontrakt så fogas dessa samman med tidigare filer i ärendet. Det innebär att filerna

som lagrats på disk kan innehålla både äldre kontrakt, gamla uppgifter om uppdateringar och aktuell information. Filerna är dock strukturerade så att nyare tillägg sparas sist i dokumenten. Totalt finns cirka 30 000 kontraktfiler.

Det finns vissa standardmallar för hyreskontrakten. Dessa används dock inte i alla fall. De har också förändrats över tid. Men ofta finns en förstasida i kontraktet med relativt tydlig struktur och visuellt stöd (i form av exempelvis rutor) för att underlätta ifyllning och läsning. I nästan alla fall är kontrakten ifyllda maskinellt, men handskrivna uppgifter förekommer (främst vid rättningar). Vissa uppgifter som sparas i KAR kommer dock inte från kontraktssidan, utan kan ligga i bilagorna eller i sidor som skrivits som löpande text.

Figur 2: Första sidan av ett avtal

Faberge		HYRESKONTRAKT FÖR LOKAL		Sid 1 (4)	
Undertecknande parter har denna dag träffat följande hyresavtal.			Nr: 5410-1012,4		
Kryss i ruta innebär att den därefter följande texten gäller					
Hyresvärd	Namn: Faberge Storstockholm AB		Personnr/orgnr: 556130-9039		
Hyresgäst	Namn: Ekonomistyrningsverket (ESV)		Personnr/orgnr: 202100-5026		
	Aviseringsadress: Box 45316, 104 30 Stockholm				
Lokalens adress m.m	Kommun: Stockholm		Fastighetsbeteckning: Kv Barnhusväderkvarnen 36		
	Gata: Drottninggatan 89		Trappor/hus: 1 och 2 tr	Lokalens nr:	
Lokalens användning	Lokalen med tillhörande utrymnen hyrs ut för att användas till Kontor <input type="checkbox"/> Användningsändamålet har närmare angetts i bifogad specifikation. Bilaga:				
Lokalens skick	Lokalen hyrs ut i befintligt skick. <input type="checkbox"/> En beskrivning av lokalens skick vid detta kontrakts tecknande samt uppgift om vem som på tillträdesdagen i förekommande fall ska ha ombesörjt och bekostat dels åtgärdandet av brister, dels överenskomna ändringar framgår av bifogad besiktnings- och åtgärdsprotokoll. Bilaga:				
Lokalens storlek och omfattning	Areatyp	Plan	ca m ²	Areatyp	ca m ²
	Kontorsarea	6	953	Kontorsarea	1765
Om i kontraktet angiven area avviker från faktisk area medför avvikelser inte rätt för hyresgästen till återbetalning eller sänkning av hyran respektive hyresvärdens rätt till högre hyra. <input checked="" type="checkbox"/> Omfattningen av den förhyrda lokalen framgår av bifogade ritning(ar). Bilaga: 2					
<input type="checkbox"/> Tillfart för bil för i- och urlastning <input type="checkbox"/> Plats för skylt <input type="checkbox"/> Plats för skyltskåp/automat <input type="checkbox"/> Parkeringsplats(er) för _____ bil(ar) <input type="checkbox"/> Garageplats(er) för _____ bil(ar) <input type="checkbox"/>					
Hyrestid	Från och med den: 2020-02-01		Till och med den: 2025-01-31		
Uppsägningstid/ Förlängningstid	Uppsägning av detta kontrakt ska ske skriftligen minst 12 månader före den avtalade hyrestidens utgång. I annat fall är kontraktet för varje gång förlängt med <input checked="" type="checkbox"/> 3 år <input type="checkbox"/> _____ månader				

Filerna håller varierande digital kvalitet. Vissa är skannade, skeva och har en låg upplösning. Andra är digitalt signerade och har en mycket hög upplösning.

Figur 3: Ett tidigare avtal för samma aktörer som i figur 2, med handskrivna fält och skevt inskannat

Fabege		HYRESKONTRAKT För i första hand lager, förråd och lokaler av motsvarande karaktär	
Undertecknade har denna dag träffat följande hyresavtal:		Kryss i ruta innebär att den därefter följande texten gäller.	
Hyresvärd	Fabege Storstockholm AB	Personnr/orgnr	556130-9039
Hyresgäst	Ekonomistyrningsverket (ESV)	Personnr/orgnr	202100-5026
Lokalens adress m m	Kommun Stockholm	Fastighetsbeteckning Barnhusväderkvarnen 36	Lägenhet nr
	Gata Drottninggatan 89		
	Aviseringsadress Box 45316, 104 30 Stockholm		
Lokalens skick	Lokalen med tillhörande utrymnen uthvrs. om inte annat annes. I bifäst fält ska det användas till: T. 1. 1. 1.		

Vissa uppgifter skrivs av rakt från kontrakten in i KAR. Men vissa fält, exempelvis hyresarean, beräknas av handläggaren genom att summera arean för olika våningsplan. Andra sparas med en annan formulering (t.ex. "Ekonomistyrningsverket (ESV)" blir "Ekonomistyrningsverket" i avtalet som står i figur 1 och 2) eller blir sammanslagna ("Fabege Storstockholm AB" och andra regionala bolag blir "Fabege AB"). I dessa fall finns ingen exakt överensstämmelse mellan uppgiften i KAR och uppgifterna i filen.

I rapportens underrubrik har vi beskrivit dessa dokument som "ostrukturerade" och det går att nyansera när man ser tabellerna i figur 2 och 3. Men många hyresavtal följer inte dessa mallar och även de som gör det visar ofta en unik variation som gör det väldigt svårt att extrahera uppgifter med en regelbaserad metod.

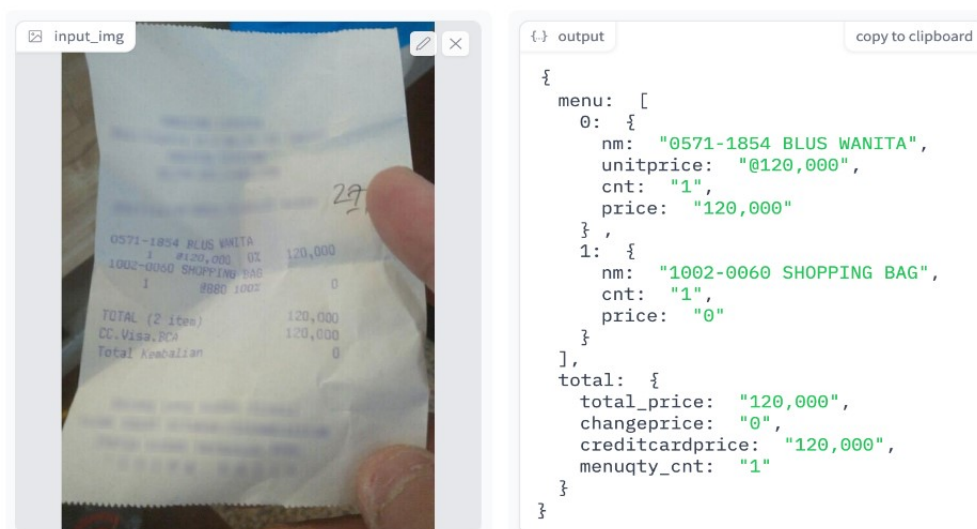
2 Beskrivning av modellerna

2.1 Donut

Donut är en ny metod för att extrahera ostrukturerade uppgifter från dokument. Förkortningen Donut står för *Document understanding transformer*¹. Som namnet avslöjar bygger metoden på en Transformers-baserad modell med en bildkodare (*visual encoder*) och en textavkodare (*textual decoder*). Modellen har tränats på en större mängd annoterade dokument och måste finjusteras vidare på en annoterad datamängd för att leverera goda resultat inom det område den används.

Enligt forskarna som har utvecklat metoden kan den uppnå ledande prestanda för olika typer av dokument (kvitton, fakturor, m.m.). Donut kan dessutom användas för olika typer av arbetsuppgifter: utöver uppgiftsextrahering kan den även användas för att klassificera dokument och svara på frågor om dokumentets innehåll.

Figur 4: Bild från Donuts presentation, digitalisering av kvitto



Eftersom Donut inte bygger på OCR-baserad extrahering behöver den inte tränas på en annoterad datamängd där uppgifternas platser på sidan har kodats, vilket var en fördel för oss i denna studie. Men den behöver finjusteras av användaren på en träningsmängd för att ge bäst resultat.

¹ Geewook, Kim m.fl. (2022), [OCR-free Document Understanding Transformer](#) (v5). arXiv:2111.15664.

2.2 GPT-4-Vision

GPT-4-Vision är namnet på en ny kommersiell produkt som företaget OpenAI lanserade på sin första utvecklarkonferens (den s.k. *OpenAI DevDay*²). Produkten är tillgänglig som en molntjänst och kostar några kronor per användning. Eftersom OpenAI inte i detalj beskriver tekniken som ligger bakom sina produkter är det oklart exakt hur den fungerar. Men tjänsten ger den stora språkmodellen GPT-4-Turbo kapaciteten att ta emot en eller flera bilder och besvara frågor om dem. Sådana modeller kallas ibland för stora multimodala språkmodeller och fältet utvecklas snabbt (se kapitel 5).

Figur 5: Exempel av GPT-4-Vision, OpenAI:s GPT-4 Technical Report³



GPT-4 The image shows a package for a "Lightning Cable" adapter with three panels.

Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.

Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.

Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

Produkten kan i nuläget inte finjusteras (*fine-tuning*), till skillnad från andra produkter från OpenAI, så det är med den skriftliga instruktionen som kommer med bilden (*prompt*) som användaren kan påverka resultatet.

² Open AI, *New models and developer products announced at DevDay*, <https://openai.com/blog/new-models-and-developer-products-announced-at-devday>. Publicerad 6 november 2023. Hämtad 1 mars 2024.

³ OpenAI m.fl. (2023), [GPT-4 Technical report](#) (v4). arXiv:2303.08774

3 Metodbeskrivning

För de två modellerna, Donut och GPT-4-Vision, har vi velat utveckla ett dataflöde i syfte att kunna testa resultaten på ett flertal datafält, för att se hur extraheringens kvalitet skiljer sig åt mellan olika typer av uppgifter (datum, nyckeltal såsom hyra).

Vi testade följande fält (numrering i Figur 6):

- Hyresvärdens namn (1)
- Hyresgästens namn (2)
- Hyresgästens organisationsnummer (3)
- Fastighetens fastighetsbeteckning (4)
- Fastighetens adress (5)
- Fastighetens kommun (6)
- Fastighetens area (7)
- Fastighetens hyra (saknas från figuren)
- Avtalets startdatum (8)
- Avtalets slutdatum (9)

Figur 6: Samma avtal med de testade fälten

Sid 1 (4)

Fabege HYRESKONTRAKT FÖR LOKAL Nr: 5410-1012,4

Undertecknande parter har denna dag träffat följande hyresavtal. Kryss i ruta innebär att den därefter följande texten gäller.

Hyresvärd	Namn: Fabege Storstockholm AB 1		Personnr/orgnr: 556130-9039
Hyresgäst	Namn: Ekonomistyrningsverket (ESV) 2		Personnr/orgnr: 202100-5026 3
Lokalens adress m.m.	Kommun: Stockholm 6	Fastighetsbeteckning: Kv Barnhusväderkvarnen 36 4	
	Gata: Drottninggatan 89 5	Trappor/hus: 1 och 2 tr	Lokalens nr:
Lokalens användning	Lokalen med tillhörande utrymmen hyrs ut för att användas till Kontor <input type="checkbox"/> Användningsändamålet har närmare angetts i bifogad specifikation. Bilaga:		
Lokalens skick	Lokalen hyrs ut i befintligt skick. <input type="checkbox"/> En beskrivning av lokalens skick vid detta kontrakts tecknande samt uppgift om vem som på tillträdesdagen i förekommande fall ska ha ombesörjt och bekostat dels åtgärdandet av brister, dels överenskomna ändringar framgår av bifogad besiktnings- och åtgärdsprotokoll. Bilaga:		
Lokalens storlek och omfattning	Areatyp	Plan	ca m ²
	Kontorsarea	6	953
	Areatyp	Plan	ca m ²
	Kontorsarea	7	1765 7
	Om i kontraktet angiven area avviker från faktisk area medför avviken inte rätt för hyresgästen till återbetalning eller sänkning av hyran respektive hyresvärdens rätt till högre hyra. <input checked="" type="checkbox"/> Omfattningen av den förhyrda lokalen framgår av bifogade ritning(ar). Bilaga: 2		
	<input type="checkbox"/> Tillfart för bil för i- och urlastning	<input type="checkbox"/> Plats för skylt	<input type="checkbox"/> Plats för skyltskåp/automat
	<input type="checkbox"/> Parkeringsplats(er) för bil(ar)	<input type="checkbox"/> Garageplats(er) för bil(ar)	<input type="checkbox"/>
Hyrestid	Från och med den: 2020-02-01 8		Till och med den: 2025-01-31 9

Fälten valdes eftersom de oftast finns på avtalens första sida men också för att de utmanar modellernas förståelse av uppgifterna på olika sätt:

- Hyresgästens organisationsnummer är ofta i närheten av hyresvärdets organisationsnummer.
- Fastighetsbeteckning finns inte alltid och kallas även för ”Kvarter” i vissa avtal.
- Fastighetens area anges i olika former, ibland som ”area om” (ovan mark), ibland olika våningars värden summeras för att få den totala arean.
- Fastighetens hyra sparas som årshyra i KAR men kan anges per månad eller kvartal i avtalen.
- Avtalets datum kan anges i olika format eller ibland som ”tillsvidare”.

Vi testade också med ett större urval av avtal ur datamängden för att både träna och utvärdera modellerna på en mängd olika utmaningar. Trots att avtalens kvalitet blivit högre över tid ville vi kunna se hur modellerna fungerade med sämre dokument.

3.1 Framtagning av en test-datamängd

Som tidigare nämnts är databasens innehåll inte exakt överensstämmande med avtalens uppgifter. Handläggarna summerar värden, expanderar förkortningar och kompletterar uppgifter som saknas.

Vi ville främst testa modellernas kapacitet att hitta informationen och att extrahera den. Vi bestämde oss därför tidigt för att välja ett urval av avtal och manuellt annotera dem. Detta gjorde vi med ungefär 180 avtal, varav 100 skulle utgöra vårt test-urval.

För att kunna skicka avtalsdokumenten som bild till OpenAI:s servrar maskerade vi alla personuppgifter.

3.2 Donuts finjustering

För att finjustera Donut-modellen använde vi 1 000 avtal från datamängden. De annoterade uppgifterna kom direkt från KAR med de ovannämnda avvikelserna. Finslipningsprocessen bestod av ett flertal omgångar (s.k. *epochs*) och kördes med 80 procent av datamängden. Efter varje omgång validerades den med hjälp av 10 procent av datamängden som inte användes för finslipningen. Till slut utvärderade vi modellens prestanda på de sista 10 procenten, och även med de 100 avtalen i test-urvalet.

Att finjustera modellen tog ungefär 2 timmar på datalabbets arbetsstation. Processen avslutades automatiskt efter 3 pass (*epochs*) när valideringsarbetet slutade att visa framsteg.

3.3 Instruktionsoptimering för GPT-4

Den andra metoden vi använde byggde på modellen GPT-4-Vision. För att optimera resultatet var det inte möjligt att finjustera modellen. Däremot kunde vi ge modellen mer kontext och mer detaljerade instruktioner. Detta kallas på engelska *prompt engineering* och har blivit ett stort forskningsområde sedan ChatGPT lanserades.

I vårt fall kunde vi till exempel få resultat i JSON-format genom att specificera det till modellen med formuleringen ”Ditt svar är i JSON-format”, även om den ibland valde att komplettera eller ersätta det med en vanlig mening på svenska. Vi styrde också formatet för vissa fält med särskilda instruktioner.

Instruktionen som slutligen användes var denna:

Du är en svensk handläggare på en myndighet som extraherar nyckelinformation från hyresavtal.

Du ska hitta de följande uppgifterna: { 'hyresvärd', 'hyresgäst', 'fastighetsbeteckning', 'adress', 'kommun', 'orgnr', 'hyra', 'area', 'startdatum', 'slutdatum' }.

Du bör följa de följande riktlinjerna:

- *Ditt svar är i JSON-format med tidigare nämnda variabelnamn och innehåller BARA de efterfrågade uppgifterna.*
- *Om en uppgift saknas så ange NA.*
- *orgnr avser hyresgästens organisationsnummer, INTE hyresvärdens orgnr.*
- *adress avser hyresobjektets adress, inte hyresgästen adress och inte hyresvärdens adress.*
- *adress ska endast innehålla gata och gatunummer, inte postnummer och inte postadress.*
- *datum ska vara formaterade enligt YYYY-MM-DD.*
- *hyra innehåller endast siffror.*
- *area innehåller endast siffror.*

Var extra noggrant för att undvika misstag eller typo.

Kan du extrahera information från detta avtal?

4 Resultat

Resultaten redovisas på följande sätt. För varje fält som transkriberas klassificerar vi resultatet i en av de följande kategorierna:

- **Perfekt match:** texten är korrekt och bör kunna läggas till i databasen utan större omformulering.
- **Annan formulering:** texten är korrekt i sak men en handläggare hade formulerat den annorlunda (t.ex. Stockholms kommun Fastighetsförvaltningen i stället för Stockholms kommun för en hyresvärd)
- **Stavning:** texten är så pass nära en perfekt match att det är uppenbart att den plockades från rätt fält men stavningen är fel (t.ex. Vidjan 19 istället för Vidåsen 19)
- **Fel:** texten har lite att göra med det som finns i dokumentet, vilket tyder på att det inte hittades eller transkriberades med stora stavningsfel (t.ex. Skolgatan 29 i stället för Järnvägsgatan 30).

I kategorin Fel har vi även inkluderat alla tillfällen där uppgiften inte hittades eller där en uppgift ”hittades” trots att den inte fanns på sidan.

Observera att ett stavfel kan ha olika betydelser för olika fält. För en adress kan det vara lätt att upptäcka (t.ex. Skolläraren 2 i stället för Skolvägen 2), ibland kan det till och med fixas automatiskt (det skulle vara enkelt med kommuner där möjligheterna är få och kända). Men för ett numeriskt värde såsom en hyra, ett datum eller en area (t.ex. 13 000 i stället för 1 300) är konsekvenserna större och möjligheterna att upptäcka felet mindre. Detta skiljer sig dock inte från mänskliga stavfel.

4.1 Donut

Att maskinellt extrahera informationen ur 100 dokument tog 62 sekunder på datalabbets arbetsstation, mindre än en sekund per dokument.

Tabell 1: Transkriberingsprestanda med Donut-modellen

Fält	Perfekt match	Annan formulering	Stavning	Fel
Slutdatum	91	-	3	6
Startdatum	86	-	2	12
Hyresvärd	27	25	26	22
Fastighet	39	1	37	23
Kommun	69	-	7	24
Adress	37	-	37	26
Area	59	4	9	28
Hyresgäst	43	25	4	28
Hyra	54	-	8	38
Genomsnitt	56 %	6 %	15 %	23 %

För Donut-modellen är 56 procent av fälten perfekt extraherade. I ytterligare 6 procent har modellen angett en alternativ formulering. Detta är oftast en konsekvens av vår finjustering som bygger på data som handläggarna har ändrat och inte nödvändigtvis negativt (t.ex. ”Fabege Storstockholm AB” till ”Fabege AB”). I 15 procent har modellen framgångsrikt hittat fältet men transkriberat dess innehåll med små stavfel. I 23 procent är texten helt fel transkriberad.

Det finns dock stora skillnader mellan fälten. Datum transkriberas exakt i över 86 procent av avtalen, och *Area* i 59 procent av dem. Fältet *Kommun* blir rätt i 69 procent av fallen. *Hyresvärd* och *Hyresgäst* lider av väldigt många alternativa formuleringar (25 % vardera) och en del stavfel (26 % för hyresvärden). Detta skulle kunna förklaras av att dessa två fält visar störst avvikelse mellan det som står på dokumenten och det som står i KAR. Modellen har helt enkelt tränats på fel data.

Fälten *Hyra* och *Area* visar stora felnivåer (38 % respektive 28 %), vilket delvis handlar om att Donut har svårt att upptäcka när informationen inte finns och att dessa två typer av uppgift saknas oftare än de andra på sidorna vi har testat med. Där är det också kopplat till kvalitetsproblem i vår datamängd eftersom värdena alltid finns i KAR, men inte alltid just på sidan vi ger till modellen. I de fall uppgiften saknades på sidan upptäckte modellen detta endast 9 av 75 gånger (12 %).

4.1 GPT-4-Vision

Att transkribera 100 dokument med GPT-4-Vision tog 18 minuter och 15 sekunder, ungefär 11 sekunder per dokument.

Tabell 2: Transkriberingsprestanda med GPT-4-Vision

Fält	Perfekt match	Annan formulering	Stavning	Fel
Startdatum	90	-	6	4
Slutdatum	88	-	6	6
Hyresgäst	82	6	2	10
Hyresvärd	72	11	4	13
Hyra	73	-	13	14
Adress	69	2	14	15
Kommun	83	-	1	16
Fastighet	57	2	22	19
Area	69	1	7	23
OrgNr*	64	-	4	32
Genomsnitt	75 %	2 %	8 %	15 %

*OrgNr avser hyresgästens organisationsnummer, hyresvärdens sparas inte i KAR.

GPT-4-Vision klarar sig betydligt bättre än Donut-metoden, med 75 procent som transkriberas exakt och enbart 15 procent helt felaktiga resultat.

Tjänsten är bra på att hitta datum och transkribera dem (ungefär 90 % exakta svar). Den är däremot sämre på att hitta värden som *Hyra* (73 %) och *Area* (69 %), vilket kan förklaras med att modellen har svårt att transkribera vissa siffror. Den glömmer ibland en av dem (t.ex. 49,5 i stället för 249,5) eller lägger till en felaktig siffra (1249,5 i stället för 249,5).

Till skillnad från Donut stavar GPT-4 nästan aldrig fältet *Kommun* fel (det enda tillfället var Vaggeryd istället för Vaggeryd). Det tyder på att den bakomliggande modellen har en avancerad förståelse för det svenska språket och kan gissa rätt när texten är svårläst i avtalet, på samma sätt som en människa skulle göra. Både *Fastighet* och *Adress* lider av fler stavningsfel.

GPT-4-Vision är mycket bättre än Donut på att identifiera när uppgiften inte finns i avtalet (51 av 96 stycken, eller 53 %) men det är fortfarande en av de svåraste arbetsuppgifterna för tjänsten. Även att identifiera rätt organisationsnummer för hyresgästen när uppgiften saknas eller blandas med hyresvärdens nummer var mycket svårt (32 % fel).

Tabell 3: Felnivåer för Donut och GPT-4-Vision

Fält	Donut	GPT-4-Vision	Skillnad
Slutdatum	6	6	0
Startdatum	12	4	8
Hysesvärd	22	6	16
Fastighet	23	19	4
Kommun	24	16	8
Adress	26	15	11
Area	28	23	5
Hysesgäst	28	6	22
Hyra	38	14	24
Genomsnitt	23 %	15 %	11 %

Överlag visar GPT-4-Vision mycket lägre felnivå samt mycket högre nivå av perfekta matcher. En del av skillnaden kan sannolikt förklaras med avvikelserna i datamängden som vi använde för att finjustera Donut-modellen, men det är samtidigt till OpenAI:s förtjänst att det uppnådde en så hög nivå utan att få se något exempel i förhand. GPT-4-Vision klarade sig betydligt bättre på ”udda” dokument som inte följde en typisk avtalsmall.

5 En snabb teknisk utveckling

Vi började detta projekt i oktober 2023 och skrev denna rapport under vintern 2024. Under perioden har automatisk extrahering av information från dokument varit ett väldigt hett ämne med nyheter nästan varje vecka. Största nyheten kom 6 november när OpenAI släppte tillgång till ett API till GPT-4-Vision⁴. Vi bestämde oss då för att inkludera den i vår jämförelse (och exkluderade andra modeller ur studien).

Men utvecklingen av multimodala modeller fortsatte under hösten med flera, även öppna, modeller som visade lovande resultat, såsom Llava⁵. Även Google presenterade sina nya Gemini-modeller den 6 december 2023 med stort fokus på multimodalitet⁶. Vissa mer specialiserade lösningar såsom DocLLM⁷ utgår från dokumentens struktur i stället för att se deras sidor som bilder. Vi har inte hunnit utvärdera dessa alternativ och valde att fokusera på GPT-4-Vision, men det finns initiativ för att sammanställa kunskapsläget och mäta olika modellers resultat⁸.

Parallellt med detta fortsatte ”traditionella” modeller (som bara läser text) att visa intressanta resultat för att strukturera ostrukturerad information. Vi har också provat dem men inte utvärderat ordentligt inom detta projekt. Däremot ligger det nära några av datalabbets andra projekt där vi försöker att skapa assistenter som kan stötta ESV:s medarbetare genom att svara på frågor och formulera texter utifrån medarbetarnas kontext. En förutsättning är att assistenten kan ta emot dokument från handläggaren och förstå dem. Dessutom fortsätter utvecklingen av lösningar som inte använder stora språkmodeller, utöver Donut. Bland de andra kan exempelvis LayoutLM⁹ nämnas.

⁴ Open AI, *New models and developer products announced at DevDay*, <https://openai.com/blog/new-models-and-developer-products-announced-at-devday>. Publicerad 6 november 2023. Hämtad 1 mars 2024.

⁵ Yanzhe, Zhang .fl. (2024), *LLaVAR: Enhanced Visual Instruction Tuning for Text-Rich Image Understanding* (v2). arXiv:2306.17107.

⁶ Google, *Introducing Gemini: our largest and most capable AI model*, <https://blog.google/technology/ai/google-gemini-ai/>. Publicerad 6 december 2023. Hämtad 1 mars 2024.

⁷ Dongsheng, Wang m.fl.(2023) *DocLLM: A layout-aware generative language model for multimodal document understanding* (v1). arXiv:2401.00908.

⁸ Chaoyou Fu m.fl., *Awesome-Multimodal-Large-Language-Models*, <https://github.com/BradyFU/Awesome-Multimodal-Large-Language-Models/>. Hämtad 1 mars 2024.

⁹ Yiheng, Xu m.fl. (2020), *LayoutLM: Pre-training of Text and Layout for Document Image Understanding* (v5). arXiv:1912.13318.

Samtidigt som vi ser ny teknik utvecklas är det fortfarande få lösningar som använder den. Just nu visar vissa verktyg, såsom ChatGPT, goda resultat men med ett gränssnitt och med användarvillkor som inte passar att använda i KAR:s processer. De mer specialiserade tjänsteleverantörerna vars produkter vore enklare att upphandla och implementera i verksamheten har ännu inte implementerat de nya mer kraftfulla modellerna i sina produkter. Därför kommer 2024 bli ett intressant år där AI blir en mer närvarande del av många kontors-it, exempelvis genom funktioner såsom Microsoft Copilot i Windows och Office¹⁰.

¹⁰ Computer Sweden, *Uddevalla kommun bland de första med 365 Copilot – "känns coolt"*, <https://computersweden.se/article/1291650/uddevalla-kommun-bland-de-forsta-med-365-copilot-kanns-coolt.html>. Publicerad 5 december 2023. Hämtad 1 mars 2024.

6 Slutsatser

Våra tester med både Donut och GPT-4-Vision visar att det går att automatisera extraheringen av uppgifter från myndigheternas hyresavtal med en ganska hög nivå av korrekthet. Maskinen gör fortfarande ett stort antal misstag, så handläggaren behöver fortfarande kontrollera och komplettera uppgifterna, men en sådan funktion integrerad i arbetsflödet (dvs. i deras arbetsverktyg KAR) skulle kunna spara tid.

Om ett system med automatisk extrahering av information implementeras kan systemet förbättras på följande sätt:

- Integrera systemet med olika officiella register (exempelvis myndighets- och fastighetsregistret) för att automatiskt identifiera och korrigera uppenbara fel.
- Införa mekanismer för att flagga stora avvikelser i data, som orealistiskt höga hyror eller ogiltiga datum.

Testerna visar också på utmaningarna som uppstår när man försöker automatisera delar av komplicerade processer som har väletablerade rutiner. Insamlingen av avtalsuppgifterna i KAR är mer än en enkel transkriberingsprocess: handläggarna gör val, och de måste hantera dokument av varierande kvalitet. Samtidigt sparas inte uppgifterna alltid på lämpligast sätt för att vidareanvända dem till att träna en AI-modell. Några små ändringar i arbetsprocessen skulle kunna göra datamängden betydligt mer användbar i detta syfte.

Det finns vidare nya aspekter att ta hänsyn till när maskinen tar ett större ansvar i arbetet. Artificiell intelligens kallas ofta för en "svart låda" och skapar utmaningar när det gäller förklarbarhet av de beslut som tas. Processen måste därför vara helt transparent och kontinuerligt kvalitetssäkras för att minimera felaktigheter. Inkludering av medarbetare i implementeringsprocessen är avgörande för att säkerställa att tekniken möter deras behov och maximerar nytta. Även behandlingen av känsliga uppgifter, särskilt vid användning av externa tjänster, kräver noggrann uppmärksamhet för att upprätthålla dataskydd.

Detta avgränsade projekt ska också ses i en större kontext. Extraktion av uppgifter från dokument som hyresavtal är en arbetsuppgift som finns hos alla myndigheter och det finns en enorm efterfrågan för lösningar som kan underlätta eller automatisera uppgiften. Genom att samarbeta med andra aktörer och sträva efter öppna och återanvändbara lösningar kan vi maximera nyttan av våra insatser, inte bara internt utan även för den offentliga sektorn i stort. Genom att publicera denna rapport och uppmuntra till samarbete hoppas ESV kunna bidra till en bredare dialog och kunskapsutbyte inom sektorn.

ESV gör Sverige rikare

- Vi har kontroll på statens finanser, utvecklar ekonomistyrningen och granskar Sveriges EU-medel.
- Vi arbetar i nära samverkan med Regeringskansliet och myndigheterna.