



## FALLSTUDIE OM OPTIMERING OCH GENERERING AV INNEHÅLL MED AI

Ett experiment med GenAI-driven innehållsoptimering och lokalisering

FYRA TESTADE AI-VERKTYG

PROMPTER SKRIVNA OCH TESTADE PÅ  
TVÅ SPRÅK

GENOMSnittlig KVALITETSPOÄNG FÖR TYSKT INNEHÅLL OPTIMERAT MED  
ENGELSKA PROMPTER: 86 %

### BAKOM LIONBRIDGES EXPERIMENT

Lionbridge genomförde experimentet med sitt interna team och egna resurser och innehåll. Närmare bestämt använde vi vårt tyska innehåll i form av befintliga blogginlägg. Även om fallstudien fokuserade på innehåll från Lionbridge (framför allt på tyska) är avsikten att ta fram protokoll och samla kunskaper som kan användas till våra tusentals kunder inom alla sektorer och språk. Vi tänker använda resultaten och observationerna från experimentet för att hjälpa kunder med skiftande behov av skriftligt och muntligt innehåll. Vi kommer också använda informationen i vår egen marknadsföring för att utforma processer för att skapa och optimera innehåll.

### UTMANINGEN

Det kan vara både tidskrävande och kostsamt att optimera innehåll, framför allt på andra språk än engelska. Dessutom ger översättningar av befintligt innehåll inte alltid bra SEO-resultat. Ibland är de helt enkelt direktöversättningar skapade utan hänsyn till språkets inneboende nyanser eller sökord och användarbeteende och/eller information som är specifika för regionen. Generativ AI-verktyg kan hjälpa oss att optimera flerspråkigt innehåll och samtidigt ge kostnads-, tids- och arbetsbesparingar.

### BAKGRUND

De fyra **generativ AI**-verktygen Amazon Bedrock, ChatGPT (Lionbridges Azure-instans med ChatGPT 4), Google Bard och Metas Llama kan alla användas för att skapa och optimera innehåll.

De analyserar sökmönster, konkurrentinformation och användarbeteenden för att ge datadrivna förslag på sökord och innehåll.

Verktygen hjälper till att skapa optimerat och engagerande innehåll med ett högt SEO-värde, som i största möjliga mån tillgodoser användarnas förväntningar. I vissa fall kan de till och med leverera önskat innehåll bättre än en människa.

Till exempel är GenAI-verktyg ibland lika bra som människor på att följa strikta teckenbegränsningar när de skriver och översätter innehåll (i synnerhet när det gäller SEO-värde).

## LÖSNINGEN

Vi genomförde ett experiment i tre steg med GenAI-verktyg för att se vilket som gav bäst resultat vid undersökning och innehållsgenerering, i synnerhet på andra språk än engelska. Vårt team fokuserade på att lära sig bra metoder för att skriva prompter, framför allt till flerspråkigt innehåll. Vi undersökte också bra metoder för att optimera gammalt innehåll på nya språk, vilket är ett sätt att förbättra en webbplats SEO-resultat.

## METODIK

Lionbridge använde fyra AI-verktyg i undersökningen: Amazon Bedrock, ChatGPT (Lionbridges Azure-instans med ChatGPT 4), Google Bard och Metas Llama. Vi genomförde de tre stegen två gånger: ett steg för undersökning, skapande respektive analys.

**Undersökning:** Under det här steget bad vi varje AI-verktyg att utföra en sökordsundersökning och välja de två bästa frågorna för att skapa innehåll på tyska.

### Vi gav vart och ett av verktygen följande:

- En prompttyp där vi bad den agera som en tysk Content Writer och "tillhandahålla tre nya tyska sökord och tre tyska frågor".
- Källinnehåll för optimering
- Befintliga sökord

Dessutom valde vi två kompletterande frågor som vi senare skulle använda för att generera korta innehållsstycken.

Vid nyckelordsundersökningar känns det nödvändigt att ha med en människa i processen (human-in-the-loop), ett så kallat manuellt inslag. Det gäller framför allt för Bard. Verktöget gör utmärkta undersökningar, men levererar för många alternativ, så det krävs en ämnesexpert som kan välja ut nyckelord eller frågor när innehåll ska genereras.



**Skapande:** Under det här steget tog vi fram två uppsättningar prompter på tyska. Prompterna för skapande av innehåll omfattade de nyckelord som identifierats i det föregående steget. En uppsättning baserades på prompttypen Content Writer och den andra uppsättningen var avsedd för instruktionerna till uppgiften.

Vi bad en medarbetare på Lionbridges AI-avdelning att titta igenom och hjälpa till att finslipa alla prompter, så att de skulle bli så effektiva som möjligt. När prompterna var klara (se nedan i avsnittet Prompter) använde vi Lionbridges resurser för att lokalisera prompterna till tyska. Sedan körde vi de två färdiga uppsättningarna med tyska prompter genom samtliga fyra AI-verktyg och sammanställde och markerade deras föreslagna innehåll.

**Vi använde följande information för att få fram prompterna:**

- Frågor och sökord från undersökningssteget (på tyska)
- Exempel på bra innehåll som vi ville imitera (på tyska)
- Källartikeltexten (på tyska)
- Prompttyper (på engelska och tyska)
- Uppgiftsprompter (på engelska och tyska)

Eftersom vi hade två uppsättningar med prompter behövde vi köra projektet två gånger i varje AI-verktyg.

**Analys:** Under det här steget bad vi tio tysktalande affärs- och språkexperter att granska innehållet. För att säkerställa en "blind granskning" visade vi inte vilket verktyg som genererat innehållet.

**Vi lade till resultat för utvärdering i ett kalkylblad med hjälp av två bedömningsmekanismer:**

- En listruta med tre värden: Oanvändbar, Användbar efter redigering och Användbar utan redigering
- Ett fält för att poängsätta innehållet från 1 till 100

**Vi samlade in synpunkterna och granskade dem för att ta reda på följande:**

- Vilket verktyg var mest effektivt?
- Vilka metoder är bra när man ska ta fram prompter? Hur fungerar det att skapa prompter på andra språk än engelska?
- Hur bra är verktygen på att utföra en nyckelordsundersökning?
- Hur bra är verktygen på att generera frågor till innehållet?
- Hur bra är verktygen på att generera vinklar till artiklarna?
- Hur bra är verktygen på att besvara de frågor de har genererat?
- Hur kan vi uppnå optimala processer för att skapa innehåll med AI?
- Hur tar man fram effektiva prompter?

## PLANERING



Välja ut tre GenAI-verktyg att använda



Välja tre artiklar att optimera i experimentet



Hitta 10 experter till granskningspanelen



Undersöka sökord och frågor



Ta fram prompter och lokalisera dem till tyska

## EXEMPEL PÅ ANVÄNDA PROMPTER

### Prompttyper

- Du är Content Writer hos Lionbridge, du skapar innehåll relaterat till språktjänster.
- Du talar tyska och bor i Tyskland.
- Du skapar språkinnehåll på tyska avsett för tysktalande personer i Tyskland, Österrike och Schweiz ...

### Målgruppsprompter

- Personer som arbetar med marknadsföring eller kommunikation
- Personer som har en högre chefstjänst, senior manager, CEO, VP eller CMO.
- Personer som talar tyska. De bor i Tyskland, Österrike eller Schweiz ...

### Skapandeprompter

- Skapa två ytterligare textstycken för att komplettera den här artikeln ...
- Varje stycke bör vara ett svar på dessa frågor ...
- Det primära sökordet för webbadressen är ...
- De sekundära sökorden är ...

## TVÄRVETENSKAPLIGT TEAM

- 10 tysktalande prompttekniker
- 1 SEO-expert från Lionbridge
- 1 expert på digital marknadsföring från Lionbridge



## RESULTAT

### AI bör kunna bli en viktig del av alla optimeringsprojekt.

Vårt experiment bekräftade att AI-verktyg kan hjälpa till att optimera och uppdatera befintligt innehåll – även på andra språk än engelska – både smidigt och effektivt.

En lovande aspekt var att kvaliteten på resultaten förbättrades under experimentets gång. Det är avgörande att human-in-the-loop inslag finns med i alla processer för att optimera eller skapa innehåll. Människor måste medverka för att välja ut sökord till optimering bland AI-resultaten. En mänsklig granskare krävs också för att validera det slutliga innehållet. (Det kan komma att ändra sig i framtiden, eftersom kvaliteten på resultaten förbättras märkbart.)

### Vi kunde identifiera vissa bra metoder:

- För närvarande krävs manuella inslag för att validera det slutliga AI-resultatet.
- Kvaliteten på våra indata förbättrades under hela experimentet när vi lärde oss nya bra metoder.
- AI kan med större sannolikhet generera innehåll av hög kvalitet om prompterna ges i form av
  - kortare fraser
  - innehåll som enkelt kan delas upp (till exempel topp 10-listor osv.)
- För närvarande är det mest effektivt att kommunicera med AI-teknik på engelska. (Det är inte så konstigt, eftersom verktygen tränas på engelska.) Prompter på tyska resulterade inte i att AI-verktygen genererade bättre innehåll på tyska – men det är intressant att notera att verktygen producerade bättre innehåll på tyska med prompter på engelska. Obs! Det hjälper dock att ge verktyget källinnehåll, sökord, frågor och exempel på målspråket (ej engelska).
- Processer för skapande av innehåll bör använda en utförlig beskrivning av den som skapar innehållet och målgruppen. Dessutom bör man helst bygga upp promptbibliotek. På så sätt kan prompter återanvändas och enhetligt innehåll tas fram åt kunder.

**Bard tycktes ge de bästa resultaten till undersökningar.**

Resultaten var i linje med den undersökning som genererades av ett verktyg som MarketMuse, som har tagits fram specifikt för det här ändamålet. Det är värt att notera att Bards undersökningsresultat kräver manuella inslag för att välja ut godbitarna till den del av processen då innehållet skapas. Med tyska instruktioner var Bard statistiskt under de två omgångarna. Med engelska instruktioner sjönk resultatet från 82/100 till 66/100. Det här var en liten datauppsättning, så den genomsnittliga poängen för Bard (på engelska) blev 74/100.

ChatGPT:s innehåll gav något fler negativa kommentarer, men granskarna berömde ibland också dess innehåll. Meta (Llama) presterade inte lika bra och är förmodligen ännu inte redo att tillförlitligt optimera eller generera innehåll på andra språk än engelska. Kvaliteten på resultaten tycktes dock bli bättre under vårt korta projekt på en månad. Bard producerade bättre undersökningsresultat och ChatGPT blev bättre på att skapa innehåll på tyska. ChatGPT:s resultat med tyska instruktioner gav inledningsvis en poäng på 47/100.

På en enda månad steg dock resultatet till 76/100. När det gäller resultat producerade med instruktioner på engelska blev poängen 71/100. Inom en månad hade poängen stigit till 86/100. (Den genomsnittliga poängen för engelska är 78,5/100.)

Både Bard och ChatGPT var duktiga på att producera korta innehållsstycken (t.ex. bara en eller två meningar). Det kom väldigt få negativa synpunkter på deras korta innehållsstycken.

**Poängen nedan avser innehåll i långt format under experimentets två omgångar.**

**POÄNG OMGÅNG 1**

Källtext	Oanvändbar	Användbar efter redigering	Användbar utan redigering	Genomsnittspoäng
<b>ChatGPT</b> (indata på tyska)	9	9	0	47.61
<b>Bard</b> (indata på tyska)	2	14	2	68.33
<b>ChatGPT</b> (indata på engelska)	3	9	6	71.67
<b>Bard</b> (indata på engelska)	0	9	9	82.22
<b>Meta (Llama)</b> (indata på engelska)	7	8	3	46.50
<b>SUMMA</b>	<b>21</b>	<b>49</b>	<b>20</b>	

**POÄNG OMGÅNG 2**

Källtext	Oanvändbar	Användbar efter redigering	Användbar utan redigering	Genomsnittspoäng
<b>ChatGPT</b> (indata på tyska)	1	7	0	76.25
<b>Bard</b> (indata på tyska)	2	4	2	66.25
<b>ChatGPT</b> (indata på engelska)	0	4	4	86.875
<b>Bard</b> (indata på engelska)	3	4	1	59.375
<b>SUMMA</b>	<b>5</b>	<b>19</b>	<b>7</b>	

**Obs!** Meta (Llama) togs bort under omgång 2 eftersom verktyget ännu inte är redo att optimera eller generera innehåll på andra språk än engelska.