

# KI i universitets- og høyskoleeksamener i Norge

## Executive summary

KI i eksamen handler nå mest om tre spenninger: (i) **måle gyldig læringsutbytte** når generativ KI er lett tilgjengelig, (ii) **effektivisere sensur uten å svekke rettssikkerhet**, og (iii) **kontrollere juks uten uforholdsmessig overvåkning**. Det statlige KI-utvalget for høyere utdanning anbefaler mer bruk av **kontrollerte eksamensformer** og å **supplere ikke-kontrollerte vurderinger med en kontrollert del** for å opprettholde tillit til vitnemål. <sup>1</sup>

Empirien viser: LLM-basert (rubrikkstyrt) scoring kan ligne menneskelig vurdering, men kan gi **systematisk avvik** (bl.a. "snillere" poengsetting) og ustabil feedback; fjernproktoring kan gi **ulikhets-/stress-effekter**; og KI-deteksjon av tekst er ofte **for usikkert** til sanksjonering alene. <sup>2</sup>

## Definisjoner og varianter

Med KI i eksamen menes her: **AES (automatisert tekstscoring)**, **KI-assistert sensur** (triage, rubrikkforslag, begrunnelsesutkast), **proktoring/overvåkning** (inkl. atferds-/biometri-basert "flagg"), **generativ KI i studentarbeid** (tillatt/ulovlig hjelpemiddel), og **læringsanalyse** brukt til eksamensutforming (item-analyse, prediksjon av vanskelighetsgrad/risiko). <sup>3</sup>

Åpne parametere (må spesifiseres lokalt): modelltype (LLM/ikke-LLM, lukket/åpen), eksamensform (hjemme/tilsyn/muntlig), fag (tekst vs beregning), anonymisering, proktoringstype (live/AI-flagging), og evalueringshorisont (pilot vs permanent ordning). <sup>4</sup>

## Argumenter for og mot

Dimensjon	For	Mot
Pedagogikk/ vurdering	Raskere feedback/likere rubrikkbruk; kan støtte konsistens og sortere besvarelser for mer målrettet sensur. <sup>5</sup>	Risiko for målefeil/"automation bias" og svekket begrunnelse/innsyn hvis KI påvirker hva sensor leser grundigst. <sup>6</sup>
Jus/ rettssikkerhet	Klare rammer kan gi trygg innovasjon via tidsavgrensede forsøk og dokumentert praksis. <sup>7</sup>	Eksamen/karakter er et enkeltvedtak; krav om upartisk og faglig betryggende vurdering gjør "svart boks" problematiske. <sup>8</sup>
Etikk/ personvern	Mindre behov for inngripende proktoring hvis eksamensdesign endres (kontrollert del). <sup>9</sup>	Proktoring og atferdsdeteksjon kan være sterkt inngripende; krever nødvendighet/proporsjonalitet og DPIA. <sup>10</sup>
Teknisk/drift	Skalerer ved fellesløsninger (eksamensplattform/plagiatkontroll) og kan gi felles standarder. <sup>11</sup>	Kost/kompleksitet: integrasjoner, logging, sikkerhet, modell-drift, klageprosesser og leverandørbinding. <sup>12</sup>

# Empiriske funn og bevisstyrke

Bruk	Studie (design)	Kvantitative funn	Begrensninger
AES (tekst)	Aydin 2025, 590 essays	QWK=0,72; r=0,73; overlapp 83,5% mellom menneske og KI. <sup>13</sup>	Én oppgave + ett språk/setting; rubrikk og "prompt" kan gi dataset-skjevhet. <sup>13</sup>
KI-assistert sensur (beregning)	Kortemeyer 2023 (feasibility; rubrikk 0-4, ~15 runder)	KI og manuell scoring korrelerer; KI ga i snitt <b>+0,9 poeng høyere total</b> (av maks 4) og feedback var ofte plausibel men tidvis feil/misvisende. <sup>14</sup>	Simulerte "studentløsninger" (generert), ikke autentiske; ikke tidsmåling; rask modellutvikling gir lav tidsstabilitet. <sup>15</sup>
Proktorering	Woldeab 2019 (observasjon; N=631)	Total score ~likt (155 vs 156/200), men ved høy trait-angst: 144,6 vs 162,3 ( $\approx d -0,7$ ). <sup>16</sup>	Ikke RCT; liten proktor-gruppe (n=44); kan være seleksjon/setting-effekter. <sup>16</sup>
KI-deteksjon (tekst)	Fiedler & Döpke 2025 (survey-eksperiment; 63 forelesere)	Gjenkjenning nær tilfældighet: 57% "AI-tekst" og 64% "mennesketekst"; profesjonell AI-tekst <20% korrekt klassifisert. <sup>17</sup>	Utdrag 200-300 ord; generaliserbarhet til norsk språk/eksamener uklar. <sup>17</sup>

En praktisk norsk premissfaktor er omfanget av generativ KI-bruk: Studiebarometeret 2024 fant at **72%** bruker generativ KI til å forklare tema/pensum; KI-utvalget viser til samme tall. <sup>18</sup>

```
xychart-beta
  title Utvalgte effektstørrelser (illustrativt)
  x-axis ["Proktorering (høy angst) d", "LLM-sensur: +poeng/4"]
  y-axis "størrelse" -1.0 --> 1.0
  bar [-0.7, 0.23]
```

( $d \approx -0,7$  beregnet fra rapporterte M/SD;  $0,23=0,9/4$ ). <sup>19</sup>

## Policy-implikasjoner og anbefalinger for Norge

**Eksamensdesign først, teknologi etterpå.** Følg KI-utvalgets hovedlinje: øk andelen kontrollerte vurderinger, og legg til en kontrollert del til større hjemme-/semesterarbeid for å sikre validitet uten massiv overvåkning. <sup>1</sup>

**Ikke bruk KI-detektorer som bevisgrunnlag alene.** Deteksjon er usikker og kan gi feilklassifisering; bruk heller prosess-bevis (muntlig forsvar, versjonshistorikk, oppgave-design) og tydelige regler. <sup>20</sup>

**KI i sensur bør være “human-in-the-loop” og målbart.** Start med lav-risiko: rubrikkforslag, konsistenskontroll, triage og begrunnelsesutkast—men la menneskelig sensur være sensurvedtaket, i tråd med utvalgets vurdering av dagens regelverk og krav til rettssikkerhetsgarantier. <sup>21</sup>

**Risikobasert etterlevelse (AI Act + personvern):** KI til å evaluere læringsutbytte eller å overvåke/detektere forbudt adferd under test klassifiseres som **høy-risiko** i utdanningskontekst; krev DPIA og dokumenter nødvendighet/proporsjonalitet før proktorering med KI-flagging. <sup>10</sup>

**Administrative grep:** Følg UHR <sup>22</sup> sin linje om tydelige KI-retningslinjer i fagmiljø og sensoropplæring, og sikre klar hjemmel/lokal forskrift der verktøy påvirker eksamensgjennomføring. <sup>23</sup>

Implementering	Kostdrivere	Pilot-krav (min.)
KI-assistert sensur (triage)	Opplæring, QA, logging	Blind A/B mot menneskelig sensur; måle avvik, tid, klager. <sup>5</sup>
Proktorering m/KI	Infrastruktur, DPIA, support	Ulikhetstest (angst/tilrettelegging), falsk-alarm-rate, klager. <sup>24</sup>
Plagiatkontroll	Lisens, integrasjon, forskrift	Informasjon til studenter + klar rettslig forankring lokalt. <sup>25</sup>

```

flowchart LR
  A[AI-tiltak]-->B[Vurderingskvalitet]
  A-->C[Juks/etterlevelse]
  A-->D[Tid/kost]
  C-->E[Tillit/klager]
  B-->E
  D-->E

```

```

timeline
  title Milepæler (utvalg)
  2016 : Felles plagiatkontroll (Uninett/Sikt-forløper)
  2023 : LLM i høyere utdanning; feasibility-studie LLM-sensur
  2024 : UHR-retningslinjer fusk; EU AI Act vedtatt
  2025 : KI-utvalgets notat om eksamen/sensur; Sikt nye fellestjenester

```

<sup>26</sup>

```

pie title Illustrativ kostfordeling (drift+innføring)
  "Lisens/sky/modellbruk" : 30
  "Integrasjon/ID/LMS-eksamen" : 25
  "Personvern/DPIA/sikkerhet" : 20
  "Opplæring/QA/kalibrering" : 15
  "Klage/forvaltning/support" : 10

```

Norske kostdrivere sees også i Sikt <sup>27</sup> sin anskaffelses-/innføringsramme (15,9 MNOK prosjektkost; ~640 MNOK over 8 år; betydelig integrasjon/migrasjon og personvernarbeid). <sup>28</sup>

1 3 4 5 6 7 9 21 22 **files.nettsteder.regjeringen.no**

<https://files.nettsteder.regjeringen.no/wpuploads01/sites/589/2025/12/Forelopige-vurderinger-notat-fra-utvalget-om-KI-i-hoyere-utdanning.pdf>

2 14 15 **[2304.11221] Can an AI-tool grade assignments in an introductory physics course?**

<https://ar5iv.org/pdf/2304.11221>

8 **regjeringen.no**

<https://www.regjeringen.no/contentassets/d1b4ecc81268477c8496534943e49eb3/horingsnotat-om-endringer-i-universitets-og-hoyskoleloven-studieavgift-advokatutgifter-og-klagesensur.pdf>

10 **Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)Text with EEA relevance.**

[https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL\\_202401689](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL_202401689)

11 **Innføring av nye fellestjenester for digitalt læringsmiljø**

<https://sikt.no/tiltak/innforing-av-nye-fellestjenester-digitalt-laeringsmiljo>

12 27 28 **Informasjon om anskaffelser og innføringskostnader fra 2025 – Digitalt læringsmiljø**

<https://cms.sikt.no/sikt-files/download/>

u%21aHR0cHM6Ly9pc2lrdC5zaGFyZXBvaW50LmNvbS9zaXRlcy9QdWJsaXNodG9TaWt0V2ViL1NoYXJlZCUyMERvY3VtZW50cy9UaWx0YWsvTmVzdGUIM

13 **Automated scoring in the era of artificial intelligence: An empirical study with Turkish essays - ScienceDirect**

<https://www.sciencedirect.com/science/article/pii/S0346251X25001940>

16 19 24 **Microsoft Word - Woldeab.docx**

<https://files.eric.ed.gov/fulltext/EJ1227595.pdf>

17 20 **Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses - ScienceDirect**

<https://www.sciencedirect.com/science/article/pii/S1477388025000131>

18 **RAPPORT**

[https://www.nokut.no/globalassets/studiebarometeret/2025/studiebarometeret-2024-hovedtendenser\\_1-2025.pdf?utm\\_source=chatgpt.com](https://www.nokut.no/globalassets/studiebarometeret/2025/studiebarometeret-2024-hovedtendenser_1-2025.pdf?utm_source=chatgpt.com)

23 **Veiledende retningslinjer om fusk - Universitets- og høyskolerådet**

<https://www.uhr.no/temasider/karaktersystemet/veiledende-retningslinjer-om-fusk/>

25 **Lovlig plagiatkontroll**

<https://sikt.no/aktuelt/lovlig-plagiatkontroll>

26 **cms.sikt.no**

<https://cms.sikt.no/sikt-files/download/>

u%21aHR0cHM6Ly9pc2lrdC5zaGFyZXBvaW50LmNvbS9zaXRlcy9QdWJsaXNodG9TaWt0V2ViL1NoYXJlZCUyMERvY3VtZW50cy9Qb3J0ZWYlQzMIQjhsam