

AI en deontische logica: van taalmodellen naar normatieve intelligentie

J.Konstapel, Leiden, 11-9-2026.

Inleiding

De ontwikkeling van kunstmatige intelligentie heeft een merkwaardige cirkelbeweging veroorzaakt. In de klassieke AI werd intelligentie aanvankelijk opgevat als een vorm van expliciete representatie en logisch redeneren. Regels, feiten, doelen en relaties werden symbolisch vastgelegd, waarna een machine daaruit conclusies afleidde. De opkomst van machine learning en vervolgens van grote taalmodellen heeft dat paradigma grotendeels omgekeerd. Moderne AI-systemen leren patronen uit enorme hoeveelheden data en produceren vervolgens gedrag dat vaak intelligent oogt zonder dat de onderliggende kennis en redeneerregels expliciet zijn gerepresenteerd.

Juist op het terrein van normen ontstaat daardoor een fundamenteel probleem. Een intelligent systeem moet niet alleen weten wat het geval is, maar ook wat het behoort te doen. Het verschil tussen *is* en *ought* wordt praktisch relevant zodra AI-systemen zelfstandig beslissingen nemen, handelen namens mensen, toegang hebben tot financiële of medische systemen, juridische documenten interpreteren of andere agents aansturen.

Deontische logica — de formele studie van verplichtingen, verboden, permissies en verwante normatieve begrippen — krijgt daardoor een nieuwe betekenis. Zij is niet langer uitsluitend een onderdeel van filosofische logica of juridische informatica. Zij kan uitgroeien tot een formele tussenlaag tussen menselijke normen en het probabilistische gedrag van generatieve AI.

De centrale stelling van dit essay is dat de meest waarschijnlijke ontwikkeling niet bestaat uit een keuze tussen LLM's en formele logica. De ontwikkeling gaat richting een hybride architectuur waarin een taalmodel de wereld en de menselijke normatieve taal interpreteert, terwijl een formeel normatief systeem controleert welke verplichtingen, uitzonderingen, conflicten en permissies daaruit daadwerkelijk volgen. De LLM wordt daarmee niet noodzakelijk de normatieve autoriteit; hij wordt eerder de interface en semantische compiler van een explicieter normatief systeem.

Dat onderscheid is cruciaal. Een taalmodel kan overtuigend over ethiek spreken zonder daarom betrouwbaar normatief te kunnen redeneren. De recente empirische literatuur begint precies dat onderscheid zichtbaar te maken.

1. Wat is deontische logica?

Deontische logica onderzoekt de logische eigenschappen van uitspraken over wat verplicht, toegestaan of verboden is. De klassieke operatoren worden meestal weergegeven als:

- $O(p)$: p is verplicht;
- $P(p)$: p is toegestaan;
- $F(p)$: p is verboden.

Daarmee verschuift de vraag van de gewone propositielogica — “is p waar?” — naar een andere vraag: “welke normatieve consequenties volgen wanneer p verplicht, verboden of toegestaan is?”

Deontische logica is nauw verwant aan modale logica. Waar noodzakelijkheid en mogelijkheid betrekking hebben op wat in mogelijke werelden het geval is, hebben verplichting en permissie betrekking op normatief acceptabele werelden of handelingsmogelijkheden. De Stanford Encyclopedia of Philosophy behandelt deze relatie uitgebreid en laat tevens zien waarom de klassieke standaard-deontische logica onvoldoende is voor veel praktische situaties.

Een eenvoudige regel als

Het is verboden persoonsgegevens zonder toestemming te verwerken.

kan worden weergegeven als $F(\text{verwerking-zonder-toestemming})$.

Maar echte normatieve systemen zijn aanzienlijk complexer. Een wet kan bijvoorbeeld bepalen:

Persoonsgegevens mogen zonder toestemming worden verwerkt wanneer dit noodzakelijk is om een wettelijke verplichting uit te voeren.

Hier is de norm conditioneel. De relevante structuur is niet simpelweg $F(p)$, maar ongeveer:

indien C , dan $O(p)$
indien niet- C , dan $F(p)$.

Dat soort conditionele normen vormt een van de redenen waarom de ontwikkeling van deontische logica veel verder is gegaan dan de oorspronkelijke standaardmodellen.

2. Het klassieke probleem: normen zijn geen gewone feiten

Een van de belangrijkste inzichten uit de deontische logica is dat normen zich niet eenvoudig laten behandelen als gewone proposities.

Neem:

Het is verplicht een contract te betalen.

en:

Het contract wordt niet betaald.

Daaruit volgt niet dat het eerste statement logisch onwaar is. Er is sprake van een normschending.

Dit onderscheid is essentieel voor AI. Een AI-systeem dat normen representeert moet drie verschillende zaken uit elkaar kunnen houden:

1. wat feitelijk gebeurt;
2. wat volgens een norm behoort te gebeuren;
3. of de feitelijke situatie de norm schendt.

Dat lijkt triviaal, maar het is een fundamenteel verschil tussen een database en een normatief systeem.

De klassieke problematiek rond *contrary-to-duty obligations* maakt dit nog scherper. Chisholm liet zien dat normale logische representaties problemen krijgen wanneer een secundaire verplichting ontstaat nadat een primaire verplichting is geschonden. Bijvoorbeeld:

Men behoort een belofte te houden.

Als men de belofte niet houdt, behoort men zich te verontschuldigen.

Een AI-agent moet dan niet concluderen dat de tweede norm de eerste vervangt. Het moet begrijpen dat de tweede norm conditioneel geldt *na* de overtreding van de eerste. De problemen rond dergelijke contrary-to-duty-structuren behoren tot de klassieke kernproblemen van deontische logica.

Dit is rechtstreeks relevant voor AI-agents. Zodra een systeem zelfstandig handelt, moet het niet alleen weten welke actie ideaal is, maar ook hoe het moet handelen wanneer een norm al is overtreden, wanneer twee normen botsen of wanneer een uitzondering van toepassing is.

3. Van Standard Deontic Logic naar defeasible reasoning

De klassieke Standard Deontic Logic is belangrijk als fundament, maar voor praktische AI-toepassingen is zij onvoldoende.

Werkelijke normatieve systemen zijn namelijk:

- conditioneel;
- contextafhankelijk;
- hiërarchisch;
- uitzonderingsgevoelig;
- veranderlijk;
- vaak intern conflicterend.

Een verkeersregel kan bijvoorbeeld zeggen dat men niet harder dan 50 km/u mag rijden, terwijl een andere regel een ambulance onder bepaalde omstandigheden een uitzondering toestaat.

Het systeem moet dan niet alleen weten:

$F(\text{snelheid} > 50)$

maar ook:

indien ambulance = ja en noodsituatie = ja, dan geldt een uitzondering.

Hier komt *defeasible reasoning* in beeld. Een defeasible regel geldt normaal gesproken, maar kan door een sterkere of specifiekere regel worden teruggedrongen.

Defeasible Deontic Logic (DDL) is daarom bijzonder interessant voor AI. DDL combineert normatieve operatoren met niet-monotone redenering en een mechanisme waarmee conflicterende regels tegen elkaar kunnen worden afgewogen. Moderne implementaties gebruiken bijvoorbeeld een superioriteitsrelatie tussen regels.

Dit is een veel realistischer model van wetgeving en sociale normen dan een verzameling absoluut geldende logische axioma's.

Governatori, Rotolo en Sartor hebben deze ontwikkeling in de juridische context uitgebreid uitgewerkt. Hun overzicht laat zien hoe deontische logica, defeasible reasoning, juridische interpretatie, normatieve posities en normconflicten samenkomen in computational law.

4. Input/Output Logic: normen als transformatie van omstandigheden naar verplichtingen

Een andere cruciale ontwikkeling is Input/Output Logic van Makinson en van der Torre.

De basisintuïtie is eenvoudig:

gegeven omstandigheden X produceert een normatief systeem bepaalde outputs Y .

Een regel kan bijvoorbeeld worden weergegeven als:

$$X \rightarrow O(Y)$$

Het belangrijke punt is dat X niet noodzakelijk een logisch onderdeel van Y wordt. Het systeem functioneert eerder als een normatieve transformatie:

feitelijke situatie $\rightarrow$ normatieve consequentie.

Dit sluit opvallend goed aan bij AI-agenten.

Een agent ontvangt bijvoorbeeld:

De gebruiker heeft een afspraak geannuleerd.

Het normatieve systeem kan vervolgens produceren:

Het systeem moet de reservering annuleren.

Of:

Het systeem mag de vrijgekomen tijd aan een andere gebruiker aanbieden.

Input/Output Logic is juist ontwikkeld voor situaties waarin de relatie tussen input en output op een inferentieproces lijkt, maar niet dezelfde eigenschappen heeft als klassieke deductie. De oorspronkelijke literatuur noemt expliciet conditionele verplichtingen, doelen, voorkeuren, acties en beliefs.

Daarmee is Input/Output Logic conceptueel zeer dicht bij wat een normatieve AI nodig heeft.

5. De komst van LLM's verandert het probleem

De grote verandering sinds ongeveer 2020 is dat natuurlijke taal zelf een computationeel medium is geworden.

Een LLM kan uit een tekst als:

Een werkgever mag medische informatie uitsluitend verwerken wanneer dat noodzakelijk is voor een wettelijke verplichting.

redelijk goed de normatieve structuur herkennen.

Maar herkenning is niet hetzelfde als formele representatie.

Een model kan bijvoorbeeld antwoorden:

“De werkgever mag deze informatie verwerken.”

terwijl de oorspronkelijke regel slechts zegt dat verwerking *onder een specifieke voorwaarde* is toegestaan.

Het verschil tussen:

$P(A)$

en

$C \rightarrow P(A)$

is normatief fundamenteel.

Hier ligt een van de meest interessante raakvlakken tussen LLM's en deontische logica: de LLM kan functioneren als een systeem voor **norm extraction**, terwijl de formele logica verantwoordelijk wordt voor **normative inference**.

Recente experimenten bevestigen dat dit onderscheid noodzakelijk is. Onderzoek uit 2025 naar normative reasoning in LLM's testte expliciet verplichtingen en permissies vanuit zowel logisch als modaal perspectief. De onderzoekers vonden dat modellen vaak geldige redeneerpatronen volgen, maar ook specifieke inconsistenties en cognitieve biases vertonen.

Een vervolgonderzoek uit 2026 onderzocht specifiek deontisch conditioneel redeneren met Wason Selection Tasks. Daarmee wordt inmiddels niet alleen gekeken of een model het juiste antwoord geeft, maar naar het soort fouten dat het maakt.

Dit is een belangrijke verschuiving. De vraag wordt:

Kan een LLM normatieve intelligentie vertonen?

in plaats van:

Kan een LLM over ethiek praten?

Dat zijn totaal verschillende vragen.

6. De “deontic keyword bias”

Een bijzonder interessant probleem is dat taalmodellen sterk gevoelig kunnen zijn voor de oppervlaktevorm van normatieve taal.

Woorden als:

must
ought
forbidden
permitted

kunnen het model richting een normatieve interpretatie sturen, zelfs wanneer de logische structuur van de situatie dat niet ondersteunt.

Dat creëert een merkwaardige paradox.

Een LLM is juist extreem goed in taal, maar taal kan tegelijkertijd de bron van de normatieve fout zijn.

Voor formele AI is dat een serieus probleem. Een systeem dat “ought” als een semantisch signaal behandelt, kan leren dat bepaalde woorden typisch met bepaalde antwoorden samengaan. Dat is iets anders dan daadwerkelijk begrijpen waarom een verplichting volgt.

De recente literatuur over normative reasoning maakt daarom steeds sterker onderscheid tussen:

surface compliance

en:

normative competence.

Dat sluit aan bij breder onderzoek naar morele competentie van LLM's. Een recente Nature-studie waarschuwt expliciet voor het *facsimile problem*: een model kan normatief redeneren imiteren zonder dat daarmee is aangetoond dat het de relevante normatieve overwegingen werkelijk begrijpt.

7. Alignment is uiteindelijk een normatief probleem

De relatie tussen deontische logica en AI wordt nog belangrijker wanneer men AI alignment serieus neemt.

Alignment wordt vaak geformuleerd als:

Hoe zorgen we ervoor dat een AI doet wat wij willen?

Maar “wat wij willen” is zelf een normatief begrip.

Een AI moet bijvoorbeeld:

- mensen niet schaden;
- eerlijk communiceren;
- privacy respecteren;
- instructies opvolgen;
- wettelijke regels respecteren;
- niet liegen;
- in bepaalde omstandigheden juist informatie achterhouden;
- menselijke autonomie respecteren.

Deze normen kunnen met elkaar botsen.

Een gebruiker kan bijvoorbeeld vragen om vertrouwelijke informatie. De norm “help de gebruiker” botst dan met “respecteer privacy”.

Een puur gedragsmatig alignmentmechanisme kan proberen voor iedere situatie een gewenst antwoord te trainen. Maar een algemeen systeem heeft juist behoefte aan een mechanisme waarmee het kan bepalen:

1. welke normen van toepassing zijn;
2. welke normen conflicteren;
3. welke norm prioriteit heeft;
4. welke uitzondering geldt;
5. welke actie het conflict het beste oplost.

Dit is bijna letterlijk een deontisch redeneerprobleem.

Een recente analyse van “shallow alignment” stelt precies dit punt: bestaande alignmentmethoden kunnen kwetsbaar zijn omdat zij gedragsdisposities versterken zonder een robuuste capaciteit voor normatieve deliberatie te creëren. Normconflicten kunnen daardoor worden uitgebuit door adversarial prompts.

De implicatie is aanzienlijk.

Deontische logica is mogelijk geen randgebied van AI safety, maar een van de theoretische instrumenten waarmee het begrip “alignment” technisch specifieker kan worden gemaakt.

8. De toekomstige architectuur: LLM als normatieve compiler

De meest plausibele architectuur voor geavanceerde normatieve AI ziet er naar verwachting ongeveer als volgt uit.

Laag 1: perceptie en taalbegrip

Het LLM interpreteert documenten, gesprekken, wetgeving, beleidsregels en de actuele situatie.

Laag 2: norm extraction

De relevante normatieve uitspraken worden omgezet in een expliciete representatie:

$O(\text{agent}, \text{action}, \text{condition})$

$P(\text{agent}, \text{action}, \text{condition})$

$F(\text{agent}, \text{action}, \text{condition})$

met aanvullende informatie over bron, bevoegdheid, prioriteit en uitzonderingen.

Laag 3: normatieve reasoner

Een formeel systeem berekent de consequenties van deze normen.

Hier kunnen verschillende technieken worden gecombineerd:

- deontische logica;
- defeasible logic;
- Input/Output Logic;
- temporal logic;
- argumentation frameworks;
- constraint solving;
- Answer Set Programming;
- description logics;
- probabilistische modellen.

Laag 4: conflict resolution

Wanneer twee normen botsen, wordt expliciet vastgesteld:

norm A heeft prioriteit boven norm B.

of:

de situatie is normatief onbeslist.

Laag 5: action selection

Pas daarna kiest de agent een actie.

Laag 6: explanation and audit

Het systeem kan vervolgens aangeven:

Deze actie is gekozen omdat regel R1 van toepassing is, R2 door uitzondering E1 wordt gedegradeerd en de uiteindelijke verplichting A oplevert.

Dat is fundamenteel anders dan:

“Ik denk dat dit de beste oplossing is.”

Voor juridische, medische, financiële en bestuurlijke AI kan dit verschil beslissend worden.

9. LLM's kunnen bovendien de bottleneck van formele systemen oplossen

De klassieke kritiek op symbolische AI is dat mensen formele representaties moeten bouwen.

Dat is duur.

Wetboeken, contracten, beleidsdocumenten en interne bedrijfsregels bestaan in natuurlijke taal. Het handmatig omzetten daarvan in logische regels is arbeidsintensief.

LLM's bieden hier een nieuwe mogelijkheid.

Een LLM kan:

document → normatieve kandidaatregels

produceren.

Daarna kan een formele reasoner:

kandidaatregels → consistentiecontrole → formele consequenties

uitvoeren.

Recent onderzoek heeft precies deze richting onderzocht: LLM's worden gebruikt om juridische teksten om te zetten naar Defeasible Deontic Logic. De onderzoekers rapporteren veelbelovende overeenstemming tussen machinegegenereerde en door experts vervaardigde formaliseringsresultaten.

Dit kan een belangrijke technologische doorbraak worden.

De klassieke symbolic-AI-benadering had een representatieprobleem.

De LLM-benadering heeft een betrouwbaarheid- en consistentieprobleem.

De hybride aanpak probeert beide problemen tegen elkaar uit te spelen.

10. Juridische AI is waarschijnlijk het eerste grote toepassingsgebied

Het recht is bijna een ideaal toepassingsgebied voor de combinatie.

Een juridische regel is immers vaak:

indien X, dan is Y verplicht, tenzij Z.

Dat is bijna een natuurlijke taalversie van een defeasible deontische regel.

Bovendien bestaan in het recht:

- normhiërarchieën;
- uitzonderingen;
- precedentes;
- bevoegdheidsregels;
- temporele geldigheid;
- conflicterende normen;
- interpretatieregels;
- sancties;
- rechten en plichten.

Dat maakt juridische AI tot een natuurlijke experimenteeromgeving.

Recent onderzoek gebruikt bijvoorbeeld Defeasible Deontic Logic om fundamentele rechten onder de EU AI Act te modelleren. Het voorgestelde DLRisk-framework kan

rechten hiërarchisch ordenen, omstandigheden modelleren waarin een recht wordt gedegradéerd en botsingen tussen rechten en deployment-situaties expliciet maken.

Ook rule-based en case-based juridische AI ontwikkelt zich in deze richting. Onderzoek uit 2025 naar deontisch case-based reasoning richt zich expliciet op inconsistentie, incomplete kennis en de noodzaak van normatieve verklaringen voor juridische AI-systemen.

11. De EU AI Act maakt het probleem praktisch

De ontwikkeling is niet alleen academisch.

De EU AI Act introduceert verplichtingen rond hoog-risico-AI, documentatie, logging, conformiteitsbeoordeling en risicobeheersing. De actuele geconsolideerde tekst bevat bijvoorbeeld expliciete verplichtingen voor aanbieders van high-risk AI-systemen.

Daarmee ontstaat een praktische behoefte aan systemen die niet alleen zeggen:

“Dit systeem voldoet aan de regelgeving.”

maar kunnen reconstrueren:

welke norm geldt,
waarom deze norm geldt,
welke feiten relevant zijn,
welke uitzonderingen van toepassing zijn,
welke verplichting daaruit volgt,
en welk bewijs die conclusie ondersteunt.

Dat is in essentie een normatief redeneerprobleem.

Ook NIST's AI Risk Management Framework en het Generative AI Profile bewegen in de richting van systematische governance, risico-identificatie, evaluatie en documentatie.

De institutionele omgeving begint daarmee steeds meer op een normatieve computatieomgeving te lijken.

12. De grootste uitdaging: conflicterende normen

Het moeilijkste probleem zal waarschijnlijk niet zijn om:

$O(A)$

te representeren.

Het moeilijke probleem is:

$O(A)$
 $O(B)$
 $A \rightarrow \neg B$

Wat moet het systeem doen?

Dit is de situatie waarin echte menselijke normatieve systemen interessant worden. Wetgeving is geen perfect consistente axiomatische system. Bedrijfsbeleid evenmin. Mensen spreken elkaar tegen, normen veranderen en uitzonderingen worden ingevoerd.

Een goede normatieve AI moet daarom niet proberen alle inconsistenties weg te poetsen.

Het moet inconsistenties **representeren en beheren**.

Daarom zijn paraconsistente en defeasible benaderingen zo belangrijk. Een normconflict hoeft niet te betekenen dat het hele systeem logisch instort.

Recent werk over deontische closure en conflict in juridische reasoning onderzoekt bijvoorbeeld precies de vraag wat er mag worden afgeleid wanneer normen conflicteren en hoe permissions zich gedragen onder dergelijke omstandigheden.

Dit is waarschijnlijk een van de centrale theoretische problemen voor toekomstige normative AI.

13. De tweede grote uitdaging: onzekerheid

De werkelijkheid waarin AI opereert is niet volledig bekend.

Een agent weet bijvoorbeeld niet met zekerheid:

Is deze persoon werkelijk bevoegd?

of:

Is dit inderdaad een noodsituatie?

of:

Is deze toestemming geldig?

De norm kan dan volledig duidelijk zijn terwijl de feiten onzeker zijn.

Dit betekent dat toekomstige normatieve AI een koppeling nodig heeft tussen:

epistemische onzekerheid

en

deontische consequenties.

Dit is een relatief nieuw en belangrijk onderzoeksgebied. Werk uit 2026 breidt Input/Output Logic bijvoorbeeld uit naar normatief redeneren onder onzekerheid en combineert dit met voorkeuren.

Daarmee ontstaat een veel realistischer model:

waarschijnlijk feit → toepasselijke norm → waarschijnlijk normatieve
consequentie → actie onder onzekerheid.

Dat is dichterbij werkelijke intelligente besluitvorming dan klassieke binaire logica.

14. De derde uitdaging: normen veranderen

Een normatief systeem is bovendien dynamisch.

Een wet wordt gewijzigd.

Een bedrijfsregel wordt vervangen.

Een gebruiker trekt toestemming in.

Een contract loopt af.

Een veiligheidsregel krijgt tijdelijk prioriteit.

Daarom moet toekomstige normative AI niet alleen kunnen redeneren met normen, maar ook met **norm change**.

De relevante vraag is dan niet alleen:

Wat is verplicht?

maar:

Welke normen zijn op tijdstip t van kracht, welke bron heeft autoriteit en welke wijziging heeft prioriteit?

Dit brengt deontische logica samen met temporal logic, knowledge representation en versioned knowledge bases.

De toekomstige normatieve agent zal dus waarschijnlijk een soort “normatief geheugen” nodig hebben: niet simpelweg een lijst regels, maar een historisch geordend systeem van normen, bevoegdheden, uitzonderingen en wijzigingen.

15. De vierde uitdaging: wie bepaalt de normen?

Hier bevindt zich het diepste filosofische probleem.

Deontische logica kan bepalen:

Als norm N geldt, volgt actie A.

Maar de logica kan niet uit zichzelf bepalen:

Norm N behoort te gelden.

Dit is het klassieke verschil tussen normatieve representatie en normatieve legitimatie.

Een formeel systeem kan perfect consistent zijn en toch de verkeerde waarden implementeren.

Daarom is deontische AI geen oplossing voor het volledige alignmentprobleem.

De vraag:

“Wat moet de AI doen?”

kan formeel worden behandeld.

De vraag:

“Welke normen zouden voor de AI moeten gelden?”

blijft een politiek, juridisch, maatschappelijk en uiteindelijk filosofisch probleem.

Dit betekent dat normative AI waarschijnlijk drie verschillende lagen nodig heeft:

1. **norm generation** — waar komen normen vandaan?
2. **norm representation** — hoe worden ze formeel gerepresenteerd?
3. **norm reasoning** — wat volgt eruit?

De derde laag is het domein van deontische logica. De eerste twee zijn minstens zo belangrijk.

16. Van ethics naar governance

Een belangrijke toekomstige verschuiving zal daarom waarschijnlijk zijn dat men minder spreekt over “ethische AI” en meer over **governable AI**.

Ethiek is slechts één bron van normen.

Andere bronnen zijn:

- wetgeving;
- contracten;
- bedrijfsbeleid;
- gebruikersinstructies;
- veiligheidsprotocollen;
- professionele standaarden;
- technische beperkingen;
- sociale conventies.

Een AI-agent moet deze verschillende normatieve bronnen kunnen onderscheiden.

Een wet heeft bijvoorbeeld een andere autoriteit dan een gebruikersvoorkeur.

Een veiligheidsconstraint kan een gebruikersinstructie overrulen.

Een bedrijfsregel kan conflicteren met lokaal recht.

De normatieve architectuur moet dus ook **authority** modelleren.

De vraag wordt:

Wie mag wat van deze agent verlangen?

Daarmee komen deontische logica, multi-agent systems, recht en governance samen.

17. De opkomst van normative multi-agent systems

Dit wordt nog belangrijker zodra AI-systemen met elkaar gaan samenwerken.

Stel dat vijf agents gezamenlijk een taak uitvoeren.

Agent A heeft de verplichting de privacy te beschermen.

Agent B heeft de verplichting kosten te minimaliseren.

Agent C moet een deadline halen.

Agent D controleert veiligheid.

Agent E vertegenwoordigt de belangen van de klant.

Dan ontstaat een normatief systeem op organisatieniveau.

Dit is precies het soort probleem waarvoor normative multi-agent systems oorspronkelijk zijn ontwikkeld. De klassieke literatuur over normen in multi-agent systems behandelt onder meer verplichtingen, sancties, institutionele rollen en compliance.

LLM-agents maken deze oude problematiek ineens praktisch relevant.

Een organisatie van agents kan in de toekomst functioneren als een institutioneel systeem waarin:

rollen → bevoegdheden → verplichtingen → uitzonderingen → sancties

formeel worden gemodelleerd.

Dat is veel dichterbij een organisatie dan bij een chatbot.

18. De toekomst: van chatbot naar normatieve agent

De ontwikkeling van AI kan daarom worden gezien als een overgang door verschillende fasen.

De eerste generatie was:

vraag → antwoord.

De tweede generatie:

vraag → antwoord + tools.

De huidige agentische generatie:

doel → planning → acties → feedback.

De volgende stap zal waarschijnlijk zijn:

doel → normen → bevoegdheden → planning → actie → normcontrole → audit.

Daarin wordt deontische logica een soort **control plane voor agentic AI**.

De LLM genereert mogelijke acties.

De normatieve engine filtert deze:

- toegestaan
- verplicht
- verboden
- onzeker
- conflict
- menselijke goedkeuring vereist.

Pas daarna wordt een actie uitgevoerd.

Dat lijkt sterk op een compilerarchitectuur.

Natuurlijke taal is de broncode.

De LLM voert semantische analyse uit.

De deontische engine maakt formele representaties.

De reasoner berekent de normatieve consequenties.

De agent voert alleen geautoriseerde acties uit.

19. Waarschijnlijkste ontwikkelingen 2026–2030

Mijn verwachting is dat de ontwikkeling langs vijf lijnen zal verlopen.

Ten eerste zullen **LLM-based norm extraction systems** snel beter worden. Het omzetten van wetgeving, contracten en beleidsregels naar formele representaties zal waarschijnlijk een van de eerste commercieel bruikbare toepassingen worden.

Ten tweede zullen **hybride neuro-symbolische systemen** belangrijker worden. De LLM is sterk in taal en context; formele logica is sterk in consistentie, verificatie en reproduceerbaarheid.

Ten derde zullen benchmarks voor normative reasoning veel belangrijker worden. De huidige resultaten laten al zien dat algemene taalvaardigheid geen betrouwbare maat is voor normatief redeneren.

Ten vierde zal **norm conflict resolution** een zelfstandig engineeringprobleem worden. Systemen zullen expliciet moeten omgaan met prioriteiten, uitzonderingen en tegenstrijdige instructies.

Ten vijfde zal regelgeving de ontwikkeling versnellen. Naarmate AI-systemen verantwoordelijkheden krijgen in gereguleerde domeinen, wordt aantoonbaarheid van compliance belangrijker. De AI Act is daar een duidelijk voorbeeld van.

20. Waarschijnlijkste ontwikkelingen 2030 en verder

Op langere termijn kan een veel fundamenteelere ontwikkeling plaatsvinden.

Een geavanceerde AI-agent kan een **normatief model van zijn omgeving** onderhouden.

Dat model bevat:

- feiten
- personen
- rollen
- bevoegdheden
- verplichtingen
- rechten
- verboden
- uitzonderingen
- onzekerheden
- voorkeuren
- doelen
- tijd
- sancties.

Daarmee ontstaat iets dat veel dichter bij een kunstmatige institutionele intelligentie staat dan bij de huidige chatbot.

De cruciale architectonische verandering is dan dat “intelligentie” niet langer alleen betekent:

- correcte voorspelling van het volgende token,

maar:

- correcte selectie van toegestane en normatief gerechtvaardigde acties binnen een veranderende institutionele wereld.

Dat is een fundamenteel ander probleem.

21. Een mogelijke nieuwe discipline

Wanneer deze ontwikkeling doorzet, ontstaat mogelijk een afzonderlijk interdisciplinair vakgebied:

Computational Normative Intelligence.

Het zou elementen combineren uit:

- deontische logica;
- modal logic;
- defeasible reasoning;
- argumentation theory;
- machine learning;
- LLMs;
- knowledge representation;
- law and AI;
- multi-agent systems;
- AI safety;
- ethics;
- governance.

Het doel zou niet simpelweg zijn om AI “ethischer” te maken.

Het doel zou zijn:

normen computationeel representeren, interpreteren, afleiden, conflicten oplossen, uitvoeren en auditen.

Dat is een veel preciezere wetenschappelijke doelstelling.

Conclusie

De relatie tussen AI en deontische logica is inmiddels veel meer dan een theoretische curiositeit.

De klassieke deontische logica heeft een vocabulaire ontwikkeld voor verplichting, permissie en verbod. Input/Output Logic heeft laten zien hoe normen kunnen worden opgevat als transformaties van omstandigheden naar normatieve consequenties. Defeasible Deontic Logic heeft mechanismen ontwikkeld voor uitzonderingen, prioriteiten en conflicten. De juridische informatica heeft laten zien dat deze formalismen praktisch toepasbaar zijn.

LLM's brengen vervolgens een nieuwe component binnen: natuurlijke taal kan grotendeels automatisch worden omgezet in kandidaat-normatieve representaties.

Daarmee ontstaat een krachtige combinatie.

De LLM hoeft niet de volledige normatieve logica zelf te dragen.

Hij hoeft vooral te begrijpen wat mensen zeggen, welke feiten relevant zijn en welke normen vermoedelijk worden geformuleerd.

Een formele normatieve laag kan vervolgens bepalen wat daar logisch uit volgt.

Dat kan leiden tot een fundamentele verschuiving in AI-architectuur: van modellen die vooral antwoorden genereren naar systemen die expliciet kunnen onderscheiden tussen wat mogelijk, waarschijnlijk, gewenst, toegestaan, verplicht en verboden is.

De belangrijkste wetenschappelijke uitdaging ligt daarbij niet meer in het definiëren van de operator $O(p)$. Die theorie bestaat al lang.

De echte uitdaging is het bouwen van systemen die in een onvolledige, dynamische en normatief inconsistente werkelijkheid kunnen bepalen:

welke norm geldt,
voor wie zij geldt,
onder welke omstandigheden zij geldt,
welke norm voorrang heeft,
wat er gebeurt wanneer zij wordt geschonden,
hoeveel onzekerheid er over de feiten bestaat,
en welke actie uiteindelijk gerechtvaardigd is.

Daarmee krijgt de oude filosofische vraag "*wat behoort te gebeuren?*" een nieuwe technische betekenis.

Voor AI kan deontische logica uiteindelijk dezelfde rol gaan spelen die type systems, constraint systems en formele verificatie spelen in software: niet het hele intelligente systeem zijn, maar de formele laag die voorkomt dat het systeem iets doet wat het niet mag doen.

De meest waarschijnlijke toekomst is daarom geen "deontische AI" waarin een LLM wordt vervangen door logica. Het is een **hybride normatieve architectuur**, waarin statistische intelligentie en formele normatieve rede elkaar aanvullen.

Dat is waarschijnlijk de meest interessante plaats waar klassieke logica en moderne generatieve AI elkaar de komende jaren zullen ontmoeten.