

A Cat, Bought in a Bag: Consumer Transparency Challenges in AI Services

Miloš Novović, PhD¹

Abstract

Consumers who subscribe to AI services cannot find out how much access to the service they are buying. Across the industry, providers advertise ‘more usage’ and ‘five times higher limits’ – without disclosing what precisely those limits are, how consumption is measured, or when access may be reduced. Consumers thus do not know what they are paying for.

*This Article argues that EU consumer law can condemn this opacity, but cannot cure it. The transparency doctrine of the UCTD, UCPD, CRD, and DCD demands that consumers be able to evaluate the economic consequences of their commitment – a test that ‘five times’ an undisclosed, variable baseline cannot pass. Yet invalidation fixes nothing, and mandated disclosure – the *acquis*’s default remedy – fails structurally: AI subscriptions lack the measurable, predictable, stable quantity that mandatory disclosure presupposes. Token allocations are not comparable across providers, models, languages, or tasks; consumers cannot forecast their own usage – and cannot verify if that usage was proportionate to the request they have made.*

This failure marks a broader limit of disclosure-based regulation. Where consumers cannot measure or verify what they receive, the outputs are non-deterministic, and the measurements not standardized – transparency requires institutional substitution. The Article therefore proposes task-based capacity labels for consumer AI subscriptions: an independent body would measure how many representative consumer tasks – research queries, document analyses, coding sessions – each subscription tier permits and publish the results as simple comparative ratings modelled on the EU Energy Label.

¹ Associate Professor of Law, Department of Law and Governance, BI Norwegian Business School.
Email: milos.novovic@bi.no.

I. Introduction

Imagine an electricity provider which launches and heavily promotes a new, subscription-based tariff. It charges a fixed monthly fee for ‘premium energy access’ but does not state a kilowatt-hour rate anywhere. It measures consumption, instead, in its proprietary units – Happy Grids – and declines to disclose how many kilowatt-hours each entails. It does the measuring on a meter it owns, which a consumer cannot access, and which spins at different speeds at different times. The provider reserves the right to reduce supply during peak demand based on ‘grid conditions’, without specifying what such conditions are. And when a consumer contacts customer service to ask what distinguishes the premium tariff from the basic one, the representative explains that the premium access offers ‘more energy than our other plans’, refusing to specify – how much more, and what do those other plans include to begin with?

No consumer would think that entering such a contract would be a good idea. And yet, millions of European users subscribe to precisely such terms every day – when they sign up for AI services.

The pattern consistently appears across the consumer AI market, where one thing is nearly constant: a consumer cannot predict how much usage they will get out of an AI subscription they are paying for. As Part II will show, providers advertise ‘higher limits’, ‘expanded access’, or ‘more usage’ while leaving the underlying allocation undefined or variable. Consumers know the monthly price, but not the quantity of service purchased, the rate at which it will be consumed, or the extent to which it may change during performance.

EU consumer law ordinarily treats difficulties of this nature as information problems. Contract terms must permit assessment of their economic consequences; commercial practices must not omit information needed for an informed transactional decision. These rules reflect a reasonable regulatory logic: identify the missing information, require its disclosure, and allow the consumer to compare.

This Article argues that consumer AI subscriptions expose the limit of that logic.

The difficulty with AI services is not simply that providers fail to state an available figure, and should thus simply be penalised: the issue is that the service lacks a stable unit that consumers

can use across providers, predict in relation to their own use, and verify during performance. The problem is therefore not solved by replacing qualitative claims with quantitative ones: disclosure would become more exact in form while remaining unusable in substance.

The Article therefore argues that the issue is better approached as one of comparative measurement rather than informational volume. It shows why existing EU transparency doctrine can invalidate indeterminate promises yet cannot, without additional infrastructure, make it possible for consumers to meaningfully evaluate the offers on the market. It then develops a model under which an independent institution measures subscription performance through standardised consumer tasks and translates the results into a simple comparative rating. The metric suggested is *task completion capacity*: informing consumers of the number of representative tasks a subscription tier permits. The proposed approach does not require consumers to understand the internal workings of large language models; it requires the market to provide a common basis on which one subscription can be compared with another.

Part II of the Article addresses the problem of measurement: there is no standardised unit which would measure consumption and be easily communicated to consumers. Part III shows that even if a universal measurement metric could be developed, consumers would still struggle to predict the pace at which their quotas will be used. Part IV addresses the adjustment of the usage limits by the service providers. Part V proposes an institutional benchmark framework and standardised task completion capacity labels, modelled on the EU Energy Label approach. Part VI concludes.

II. The Measurement Problem

Purchasing upgraded access to an AI service is no easy task.

Anthropic's Claude Pro, billed at USD 20 per month, 'offers at least five times the usage per session compared to our free service'.² Claude Max, in turn, promises '5x or 20x more usage than Pro', with prices starting at USD 100 per month.

² Anthropic, 'Claude Plans', <https://anthropic.com/claude/plans> (accessed 8.4.2026); the quoted Pro-plan wording also at Claude Help Center, 'What is the Pro plan?', <https://support.claude.com/en/articles/8325606-what-is-the-pro-plan> (accessed 22.7.2026). The free-plan

It seems straightforward: for every single message a user on the free plan sends, a Pro user can send five, and a Max user 25 to 100.

And yet, when one tries to determine how much usage that *precisely* entails, they face a wall. The terms of use for the free service – serving as a baseline – simply state that ‘while using the free Claude plan, there is a session-based usage limit that will reset every five hours’, adding that ‘the number of messages you can send will vary based on demand, and we may impose other types of usage limits to ensure fair access to all users.’

In other words, Claude Pro subscribers pay 20 USD to receive five times more than an undisclosed, variable number; Claude Max subscribers pay in increments, to receive more increments, never knowing: the increments of what?

Anthropic is by no means alone. OpenAI’s ChatGPT Plus similarly offers ‘higher limits’ and access to ‘premium models’: which limits, and which models?³ Microsoft Copilot Pro advertises ‘more usage’ in a similar vein.⁴ A survey of six major providers’ pricing pages shows this to be a pattern: every provider substitutes qualitative vocabulary for quantitative specification – ‘Limited’, ‘Expanded’, ‘More’, ‘Higher’, ‘Highest’, ‘Unlimited’.⁵

Anecdotally, the opacity reaches the level of borderline absurdity. Users run into limits every day, without knowing when the next request will be declined – but asking the providers does not help. When Microsoft’s customer service was asked about the differences between free and paid Copilot tiers, the official response stated that, the feature having recently launched, the company

wording quoted below: Claude Help Center, ‘Get started with Claude’, <https://support.claude.com/en/articles/8114491-get-started-with-claude> (accessed 22.7.2026).

³ OpenAI, ‘ChatGPT Pricing’, <https://openai.com/chatgpt/pricing/> (accessed 8.4.2026).

⁴ Microsoft, ‘Microsoft 365 Copilot – Pricing for Individuals’, <https://www.microsoft.com/en-us/microsoft-365-copilot/pricing/individuals> (accessed 8.4.2026).

⁵ Survey conducted 7.4.2026 across pricing pages of OpenAI (n 3), Anthropic (n 2), Microsoft (n 4), Google (<https://one.google.com/about/google-ai-plans/>), xAI (<https://grok.com/plans>), and Perplexity (<https://www.perplexity.ai/pro>). All use qualitative descriptors only; xAI operates no consumer help centre.

was ‘temporarily unable to view the differences between the free and paid versions and the corresponding functional limitations in the official description’.⁶

The seller could not say what it was selling.

Microsoft is, notably, the only one of the six surveyed providers to publish a dedicated support article specifying consumer usage limits by feature and plan tier.⁷ Even there, the primary conversational feature – Chat – is listed across all tiers as ‘Extensive use’ rather than any quantified number.

What can a consumer do?

1. The Legal Framework for Measurement

The EU, as of the time of writing, does not have AI-specific contract rules: no instrument addresses the supply of AI digital services to consumers as such. The AI Act, while comprehensively regulating the use of artificial intelligence, does not address questions of a contractual nature.⁸ The Commission’s forthcoming Digital Fairness Act initiative is similarly silent in its preparatory documents.

This is not fatal on its own: where special rules do not exist, the general ones might compensate: what the EU does regulate is the contractual category into which these services fall. A paid AI subscription concluded online is a distance contract for a service – since the 2019 Modernisation Directive, expressly including a ‘digital service’ – under the Consumer Rights Directive (CRD),⁹ and a contract for the supply of a digital service under the Digital Content

⁶ Microsoft Community Support Specialist (Zenobia.Z), ‘Response to user enquiry regarding Copilot Voice free tier limits’, 11.11.2024, <https://learn.microsoft.com/en-us/answers/questions/5352475/what-are-the-free-tier-limits-for-copilot-voice-ch> (accessed 2.4.2026).

⁷ Microsoft, ‘AI credits and limits for Microsoft 365 Personal, Family, and Premium’, <https://support.microsoft.com/en-us/office/ai-credits-and-limits-for-microsoft-365-personal-family-and-premium-68530f1a-4459-4d02-9818-8233c1f673b8> (accessed 8.4.2026).

⁸ Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence [2024] OJ L, 2024/1689 (AI Act).

⁹ Directive 2011/83/EU on consumer rights [2011] OJ L 304/64, Arts. 2(6), 2(7) and 3(1), as amended by Directive (EU) 2019/2161 [2019] OJ L 328/7 (Modernisation Directive). Art. 2(16), inserted by the latter, defines ‘digital service’ by reference to Art. 2(2) of Directive (EU) 2019/770.

Directive (DCD): a service ‘that allows the consumer to create, process, store or access data in digital form’.¹⁰ The consequences of that characterisation are taken up at the end of this Section and in Part IV. Neither instrument, however, contains rules tailored to the specific difficulty this Article addresses: quantities that were stated, yet remain indeterminate.

Two instruments of general application complete the picture.

First, the Unfair Contract Terms Directive (UCTD) requires that terms offered to consumers be ‘drafted in plain, intelligible language’.¹¹ The Court of Justice has read this as a substantive, not merely formal, obligation. In *Kásler*, transparency ‘cannot therefore be reduced merely to their being formally and grammatically intelligible’;¹² the consumer must be able to ‘evaluate, on the basis of clear, intelligible criteria, the economic consequences’ flowing from the term.¹³ The standard is functional: what matters is whether the information enables assessment, not whether it parses.¹⁴ *Andriciuc* confirmed the approach for complex financial products.¹⁵

The consequences of failure are severe. Under Article 4(2) of the Directive, terms relating to the main subject matter of the contract or to the adequacy of the price escape the unfairness assessment only ‘in so far as these terms are in plain intelligible language’.¹⁶ AI subscription pricing terms – ‘5x more usage’, ‘expanded access’, ‘higher limits’ – define the main subject matter or the price/quality ratio. If they fail the transparency requirement, they lose the exemption and fall under full unfairness review under Article 3(1).¹⁷

¹⁰ Directive (EU) 2019/770 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJ L 136/1, Arts. 2(2)(a) and 3(1); recital 19 expressly includes ‘software-as-a-service’ offered in the cloud computing environment.

¹¹ Council Directive 93/13/EEC on unfair terms in consumer contracts [1993] OJ L 95/29, Art. 5.

¹² CJEU case C-26/13, *Kásler and Káslerné Rábai v OTP Jelzálogbank Zrt* ECLI:EU:C:2014:282, para. 71.

¹³ *ibid.* para. 75.

¹⁴ Marco BM Loos, ‘Crystal Clear? The Transparency Requirement in Unfair Terms Legislation’ (2023) 19 ERCL 281, 283–290.

¹⁵ CJEU case C-186/16, *Andriciuc and Others v Banca Românească SA* ECLI:EU:C:2017:703, paras. 50–51.

¹⁶ Council Directive 93/13/EEC (n 11) Art. 4(2).

¹⁷ *Kásler* (n 12) paras. 68–73.

Second, the Unfair Commercial Practices Directive complements this framework. Article 7 prohibits misleading omissions: practices that omit ‘material information that the average consumer needs, according to the context, to take an informed transactional decision’ and thereby cause, or are likely to cause, the average consumer ‘to take a transactional decision that he would not have taken otherwise’.¹⁸ For an invitation to purchase, material information expressly includes ‘the price inclusive of taxes, or where the nature of the product means that the price cannot reasonably be calculated in advance, the manner in which the price is calculated’.¹⁹

The January 2025 judgment in *Bundesverband* clarified that under Article 7(4)(c) ‘a distinction must be made between the indication of the final price and that relating to the manner in which that price is calculated’; where the final price depends on variable factors, ‘information on the application of a component ... falls within “the manner in which the price is calculated” ... and, as such, constitutes material information’.²⁰ The consumer need not be able to calculate the final price herself. What the communication must indicate is ‘the applicability in principle of such a percentage, together with a possible scale and the components having an impact on that percentage, thus enabling the average consumer to make an informed transactional decision’.²¹

Put differently: where the exact price cannot be given, disclosure of the method – its direction, range, and determinants – suffices.

Nor would formal completeness immunise the practice. In *Canal Digital Danmark*, the Court held that a communication may contain all the material information required by Article 7(4) and still constitute a misleading action or omission.²² The AI pricing page reproduces the architecture of that case precisely – a conspicuous headline figure beside inconspicuous,

¹⁸ Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market [2005] OJ L 149/22, Art. 7(1).

¹⁹ *ibid.* Art. 7(4)(c).

²⁰ CJEU case C-518/23, *Bundesverband eV v NEW Niederrhein Energie und Wasser GmbH* ECLI:EU:C:2025:35, paras. 30–31.

²¹ *ibid.* para. 52 (operative part).

²² CJEU case C-611/14, *Canal Digital Danmark A/S* ECLI:EU:C:2016:800, para. 71 and operative part, point 4.

unquantified terms – so that even a formally complete disclosure would leave the presentation itself suspect. The list can be complied with; the message can still deceive.

The Court’s recent jurisprudence extends the requirement beyond pricing mechanics to the substance of what the consumer receives. In *Justa*, transparency requires more than stating the label and amount of a fee: the consumer must understand what services the fee purchases – the counter-performance – and be in a position to verify that there is no overlap between costs or between the services they remunerate.²³ *Malicnik* confirmed: the name of a fee and the precise fixing of its amount cannot, on their own, satisfy the transparency requirement; the consumer must be able to understand the nature of the services actually provided in return.²⁴ Naming a charge is not enough; what must be explained is what the charge buys.

AI subscription pricing fails on both dimensions – the calculation limb and the counter-performance limb alike.

Consider, first, the calculation limb. *Bundesverband* does not apply directly: the headline price of a fixed-fee subscription can be calculated in advance, so the method-of-calculation limb of Article 7(4)(c) is never formally engaged; the natural home of the omission is the main-characteristics limb of Article 7(4)(a), read with Article 7(1). But the structure is the mirror image of the electricity tariff in *Bundesverband*. There, the quantum was known and a price component variable. Here, the price is fixed and the quantum variable – so that the effective price per unit of use is precisely what cannot reasonably be calculated in advance. On the logic of *Bundesverband* – where the figure cannot be given, its applicability in principle, its scale and its determinants must be – the omission fails a fortiori: not even the method is disclosed.

The counter-performance limb fares no better. The consumer does not understand what her fee purchases, a failure under *Justa*. ‘5x more usage’ specifies no rate, no unit, and no baseline: it is a ratio to an undefined quantity measured in undefined units. ‘Expanded access’ names no service. ‘Higher limits’ identifies no counter-performance; the fee is defined; what it buys is not.

²³ CJEU case C-39/24, *Justa v Banco Bilbao Vizcaya Argentaria SA* ECLI:EU:C:2025:298, paras. 39–41.

²⁴ CJEU case C-280/24, *Malicnik* ECLI:EU:C:2025:401, paras. 39–40 (reinforcing the counter-performance requirement applied in *Justa* (n 23); both rest on CJEU case C-565/21, *Caixabank (Loan arrangement fees)* ECLI:EU:C:2023:212, para. 32).

The practice leaves law in limbo. The core issue is not the *complexity* the doctrine has accommodated – the difficulty, as in *Kásler*, of calculating exchange-rate risk where the formula is disclosed but demanding. Nor is it the *uncertainty* the doctrine has addressed – the unknowability, as in *AI Telekom Austria* (under the Universal Service Directive), of future CPI movement where the indexation mechanism is transparent.²⁵ It is *indeterminacy*: the pricing equation contains an undefined term: ‘five times’ an unspecified baseline cannot be evaluated, not because the calculation is hard, but because there is nothing to calculate. The distinction is not one of degree: an indexed or even derivative obligation, however uncertain, is determinate – once the reference value materialises, the consumer’s position can be computed and verified from facts beyond either party’s control. An undisclosed baseline never settles.

The two remaining instruments – the CRD and the DCD – add no further prohibition to this indictment. They establish something more consequential: even enforcement machinery of full contractual force finds, in an indeterminate promise, nothing on which to operate.

The Consumer Rights Directive tests the defect through a third enforcement architecture. Unlike the UCTD and the UCPD, which police the quality of terms and practices, the CRD imposes affirmative duties: before the consumer is bound by a distance contract, the trader must provide, ‘in a clear and comprehensible manner’, the ‘main characteristics’ of the service and its total price – for subscriptions, ‘the total costs per billing period’.²⁶ For a fixed-fee AI subscription, the price limb is typically satisfied: the monthly charge is stated. The candidate for breach is the main-characteristics limb. The usage allocation *is* the main characteristic of a tiered subscription – and no judgment of the Court has yet construed Article 6(1)(a) for a digital subscription of undefined quantity.

The Directive’s consequences are contractual: the information provided becomes ‘an integral part’ of the contract, unalterable without express agreement;²⁷ the burden of proving compliance rests on the trader;²⁸ and the main characteristics and price must be brought to the consumer’s

²⁵ CJEU case C-326/14, *Verein für Konsumenteninformation v AI Telekom Austria AG* ECLI:EU:C:2015:782, para. 27.

²⁶ Directive 2011/83/EU (n 9) Art. 6(1)(a) and (e).

²⁷ *ibid.* Art. 6(5); CJEU case C-536/20, *‘Tiketa’ UAB v M. Š.* ECLI:EU:C:2022:112, para. 52.

²⁸ Directive 2011/83/EU (n 9) Art. 6(9).

attention ‘in a clear and prominent manner’ directly before she places the order.²⁹ Yet each of these mechanisms presupposes disclosed content on which it can operate. A binding term is useful where the term says something. ‘Expanded access’, made integral to the contract, binds the trader to nothing in particular: there is no quantum for Article 6(5) to freeze, and the burden of proof under Article 6(9) is discharged by showing that the vague text was displayed. The CRD thus completes the pattern rather than escaping it. Even the instrument with the hardest contract-law consequences has nothing determinate to grip.

The Digital Content Directive reaches the same terrain from the opposite side. It contains no pre-contractual information duties – conformity, remedies and modification are its subject matter, and disclosure is left to the CRD.³⁰ But its objective conformity standard is squarely about quantity: a digital service must ‘be of the quantity’ and possess the qualities ‘normal for ... digital services of the same type’ which the consumer ‘may reasonably expect’, taking into account any public statement made by the trader ‘particularly in advertising’.³¹ A tier marketed as ‘5x more usage’ is thus, in principle, held to the quantity such advertising makes reasonably expectable; for continuous supply, the burden of proving conformity rests on the trader.³² In practice, the conformity benchmark inherits the very indeterminacy documented above. ‘Five times’ an undisclosed baseline yields no quantum against which conformity could be assessed. And where the entire market discloses only ranked adjectives, what is ‘normal’ for services of the same type is itself unascertainable.

The failure is not for want of legal machinery – it is for want of determinable content against which the machinery can be deployed.

2. The Intuitive Remedy for Measurement Issues

The obvious response is to deem the relative terms invalid (‘five times more’) and mandate the same disclosure that AI platforms provide to developers of third-party applications which use

²⁹ *ibid.* Art. 8(2); CJEU case C-249/21, *Fuhrmann-2-GmbH v B*. ECLI:EU:C:2022:269, para. 23.

³⁰ Directive (EU) 2019/770 (n 10) Art. 1; recital 42 (the requirements of the contract include those resulting from pre-contractual information which, in accordance with Directive 2011/83/EU, forms an integral part of the contract).

³¹ *ibid.* Art. 8(1)(b).

³² *ibid.* Arts. 12(3) and 8(4).

their models. Professional developers receive API documentation that specifies pricing precisely. OpenAI publishes per-token rates differentiated by model.³³ Anthropic publishes context-window specifications and per-token prices in exact figures.³⁴ The same providers who maintain vague consumer pricing publish precise costs for parallel product lines; hence, the logical conclusion is that technical capability exists, and providers should simply publish the quantified number.

The intuitive remedy, differently stated, is to require equivalent disclosure for consumers: state that a subscription includes a defined quantity of tokens, messages, or computational resources. Two alternatives present themselves – token-based disclosure (the measure used in API pricing) and message-based disclosure (a more consumer-facing metric).

Each approach fails, but for different reasons.

3. Why Token Disclosure Fails

Tokens, the atomic word chunk units in which AI consumption is typically disclosed, fail the basic test of a measurement unit: they are not standardised.

The problem is architectural.

Different providers use different tokenisation schemes, trained on different corpora, with different compression efficiencies. Petrov and colleagues' systematic evaluation of 17 tokenisers found variation of up to 15-fold in tokenisation length for semantically equivalent text across languages.³⁵ The same sentence that costs 10 tokens at one provider may cost 150 at another – not because of pricing decisions, but because of incompatible measurement systems. A consumer told her subscription includes '1,000,000 tokens' cannot compare this to a competitor's '800,000 tokens'.

³³ OpenAI, 'API Pricing', <https://openai.com/api/pricing> (accessed 2.4.2026).

³⁴ Anthropic, 'Claude Platform', <https://claude.com/platform/api> (accessed 7.4.2026).

³⁵ Aleksandar Petrov et al., 'Language Model Tokenizers Introduce Unfairness Between Languages' (NeurIPS 2023).

Hence, at the very outset, tokens as a reliable measure fail. A kilowatt-hour represents a fixed quantity of energy regardless of supplier. A token represents a variable quantity of text depending on the provider and the model architecture.

The problem compounds across languages. Tokenisers trained predominantly on English text represent non-English languages inefficiently, producing what amounts to a language surcharge on consumption.³⁶ A query in Telugu or Amharic may require roughly four to five times the tokens of a semantically equivalent English query; even widely spoken non-Latin-script languages incur substantial multipliers.

Users in non-English markets pay more, in token terms, for equivalent service. A consumer told ‘100,000 tokens per day’ cannot translate this into ‘queries for my use case’ without knowing her language’s tokenisation efficiency relative to the provider’s training distribution.

On top of this, task types differ: code, structured data and prose tokenise differently, in ways the consumer cannot anticipate – least of all where she expects to use a model for a broad set of general, varied tasks.

Token disclosure would thus give consumers information they cannot use: formal transparency, achieved at the expense of the substantive kind. This is, ironically enough, precisely the gap *Kásler* identified – grammatical intelligibility without evaluative capacity.³⁷ The consumer learns that her subscription includes ‘100,000 tokens per day’; she does not learn, and cannot calculate, how many research queries, coding requests, or document analyses this permits relative to the alternatives at competing providers.

4. Why Message Disclosure Also Fails

If tokens are incommensurable across providers, another measure could suffice – the plain alternative being message counts. ‘300 messages per day’ has an intuitive accessibility that ‘1,000,000 tokens’ lacks. But message-based disclosure fails for different reasons; reasons that foreshadow the prediction problem addressed in the following Part.

³⁶ Orevaoghene Ahia et al., ‘Do All Languages Cost the Same? Tokenization in the Era of Commercial Language Models’ (EMNLP 2023) 9904.

³⁷ *Kásler* (n 12) paras. 71, 75.

Consumption per message, namely, is not uniform. Large language model APIs are stateless: each request must include the full conversation history for the model to maintain context.³⁸ A user's tenth message is billed not only for its own content but for the accumulated context of nine prior exchanges. The cost scales non-linearly with conversation depth. A subscription offering '300 messages' yields radically different value for brief, stateless exchanges than for extended, contextually dense conversations.

The variability compounds. 'Reasoning tokens' – internal computational steps the model performs before generating visible output – can consume up to ten times more tokens than simple inference, yet remain entirely invisible to the user.³⁹ Output tokens are typically priced two to ten times higher than input tokens, but the consumer does not control output length.⁴⁰ Two queries may look identical on the consumer's screen and differ by an order of magnitude in the resources they consume.

Even precise message counts would therefore communicate little about actual service quantum. The standard *Bundesverband* sets for a variable price component – disclosure, at minimum, of its applicability in principle, a possible scale, and 'the components having an impact' – transposed by analogy to the quantum side of the bargain, is not met by a metric that varies tenfold on factors the consumer neither controls nor observes.⁴¹

As a consequence, neither token-based nor message-based disclosures present a realistic source of transparency. The measurement problem is not an information gap that disclosure could cure: it is the absence of a common unit that would make cross-product comparison possible *in the first place*. A more attentive consumer does not overcome it. A more educated consumer does not overcome it.

³⁸ OpenAI, 'Conversation State', <https://developers.openai.com/api/docs/guides/conversation-state> (accessed 7.4.2026).

³⁹ OpenAI's reasoning documentation confirms that reasoning tokens are billed as output tokens and 'can range from a few hundred to tens of thousands' per query: OpenAI, 'Reasoning', <https://developers.openai.com/api/docs/guides/reasoning> (accessed 7.4.2026).

⁴⁰ See generally OpenAI (n 33); Anthropic (n 34).

⁴¹ *Bundesverband* (n 20) para. 52 (operative part).

Without a standardised unit – comparable across providers, stable across languages, intelligible to the consumer – the first precondition for market transparency is unmet.

III. The Prediction Problem

Let us assume the measurement problem away. Tokens are now standardised; message counts are meaningful; the consumer knows her allocation. Can she now make an informed purchasing decision?

Not yet, as she cannot anticipate the rate at which her own consumption will reach those limits.

1. Why Prediction Fails: The Technical Architecture

Four features of AI system architecture make consumption prediction structurally impossible for consumers.

The first issue is the sheer variation in *query complexity*: a simple factual question consumes fewer tokens than a nuanced analytical request. But the consumer formulating a query does not know its token cost in advance; she discovers consumption retrospectively, if at all.

The second, and more technologically challenging, issue is one of *routing*. Where providers auto-select models or offer multiple options within a subscription, the same query may consume vastly different resources depending on which model processes it. A query routed to a ‘reasoning’ model may consume orders of magnitude more than the same query processed by a standard model – yet both are presented to the consumer as the same service.

Third, *accumulation*, and fourth, *invisibility* – both already documented in Section II.4. Per-message cost accelerates with conversation depth,⁴² and the consumer cannot predict how deep a research session will run; reasoning, retrieval, safety filtering and tool use consume resources she never sees.⁴³ What Section II.4 established about the unit, this Part restates about the forecast: a consumer cannot predict a quantity whose principal drivers are invisible to her. Simply: initiating a ‘deep research’ Internet query could result in a model consulting five or five hundred webpages; a consumer does not know until she sees the output.

⁴² See n 38.

⁴³ See nn 39–40.

These are not implementation details that better-designed systems could eliminate, or attempts on the part of model providers to deceive consumers: they are inherent to the service itself.

2. The Legal Framework for Usage Prediction

The response of the legal framework appears clear. *D.V. v M.A.*, a 2023 judgment concerning Lithuanian lawyer-consumer contracts, establishes the framework for assessing transparency where a unit rate is disclosed but total cost depends on unpredictable consumption.⁴⁴

The facts were straightforward. A lawyer charged EUR 100 per hour across five contracts, specifying the hourly rate but nothing that would let the consumer approximate total cost.⁴⁵ The Court held that, absent any other information, an hourly rate alone does not enable the average consumer to estimate the total amount to be paid.⁴⁶

The structural parallel to AI service pricing is close, though inverted. In *D.V. v M.A.* the unit rate was known and the total was not: the consumer faced open-ended expenditure at EUR 100 per hour. In a consumer AI subscription the total is known and the units are not: the consumer pays a fixed monthly fee for an allocation she can neither quantify nor price. In both, one side of the exchange is unquantifiable ex ante – precisely the inability the Court held decisive.

If anything, the subscription form is the less transparent of the two. The hourly client at least knows the price of each unit she consumes. The subscriber knows neither the unit, nor the rate, nor the quantity. Only the total.

The Court acknowledged that legal services involve inherent unpredictability – court schedules, opposing party conduct, procedural complexity – but rejected it as a defence:

‘However, although a seller or supplier cannot be required to inform the consumer of the final financial consequences of his or her commitment, which depend on future events which are unpredictable and beyond the control of that seller or supplier, the fact remains that the information which the seller or supplier is required to provide before the conclusion of the contract must enable the consumer to take a prudent decision in full knowledge of the possibility

⁴⁴ CJEU case C-395/21, *D.V. v M.A.* ECLI:EU:C:2023:14.

⁴⁵ *ibid.* paras. 12–14 and 40.

⁴⁶ *ibid.* para. 40.

*that such events may occur and of the consequences which they are likely to have with regard to the duration of the provision of legal services concerned.*⁴⁷

The Court specified what this means in practice: ‘particulars that enable the consumer to assess the approximate total cost’ – such as ‘an estimate of the expected number or minimum number of hours needed to provide a certain service, or a commitment to send, at reasonable intervals, bills or periodic reports indicating the number of hours worked’.⁴⁸

Šiba, an earlier case establishing that lawyers are ‘sellers or suppliers’ under the Directive, explained why professional services require heightened transparency: ‘in the field of lawyers’ services, there is, as a general rule, some inequality between “client-consumers” and lawyers owing in particular to the asymmetry of information between the parties’.⁴⁹

If this held for lawyers’ hourly fees, it holds here with greater force: providers possess complete consumption data across millions of subscribers; consumers possess none. If hourly legal fees without cost-approximation information fail transparency, AI subscription pricing without usage-approximation information fails by the same logic – and with less justification, at that.

3. The Intuitive Remedy for Usage Prediction

The obvious response, much like the measurement problem, is mandated usage disclosure.

The basic logic is simple enough, and the narrative is easy to construct. Providers could disclose expected consumption ranges for typical use cases: ‘subscribers in your usage category typically consume X to Y messages per month.’ They do not. Providers could specify guaranteed minimum allocations: ‘this tier includes at least X messages per day before throttling may occur.’ They do not. Providers could commit to usage dashboards enabling consumers to track consumption against allocation – and some do provide consumption-to-date information – but none discloses the allocation against which consumption is measured.

These are precisely the remedies *D.V. v M.A.* contemplates. They work for legal services because legal services, despite inherent unpredictability, permit rough approximation. A lawyer

⁴⁷ *ibid.* para. 43.

⁴⁸ *ibid.* para. 44.

⁴⁹ CJEU case C-537/13, *Šiba v Devėnas* ECLI:EU:C:2015:14, paras. 23–24.

knows – and can communicate – that a contested divorce typically requires more hours than an uncontested one. The consumer receives ranges; the ranges enable prudent decisions.

Providers possess the data to generate equivalent disclosure for AI services. Usage patterns across millions of subscribers would enable estimates: ‘a casual user typically consumes X messages per month; research-intensive use requires Y; coding assistance falls in range Z.’ The information exists. The intuitive solution is to require its disclosure.

And yet: even if providers disclosed such usage ranges, the remedy would fail, for reasons both temporal and cognitive.

4. Why Usage Prediction Transparency Fails

The temporal problem is straightforward. The *D.V. v M.A.* framework requires that the consumer receive information *before* concluding the contract. Pre-contractual transparency is foundational: as the Court held in *RWE Vertrieb*, lack of information before contract conclusion ‘cannot, in principle, be compensated for by the mere fact that consumers will, during the performance of the contract, be informed’.⁵⁰

A dashboard becomes useful only *after* the consumer has subscribed – and only after she has accumulated usage data. It does not satisfy the requirement; it converts the consumer into her own analyst, generating the information the trader failed to provide. By that point, the transactional decision the transparency doctrine was meant to inform has already been made.

The deeper problem is that consumers cannot translate disclosed information into predictions about their own behaviour. Even if providers published typical usage ranges before the consumer subscribes, she could not reliably determine whether she is typical.

Behavioural economics demonstrates that prediction failure persists even for observable, controllable consumption. DellaVigna and Malmendier’s foundational study of gym membership examined consumers choosing between flat monthly fees and per-visit payment.⁵¹ Members surveyed had forecast 9.5 monthly visits; their actual attendance averaged 4.17.

⁵⁰ CJEU case C-92/11, *RWE Vertrieb AG v Verbraucherzentrale Nordrhein-Westfalen eV* ECLI:EU:C:2013:180, para. 51.

⁵¹ Stefano DellaVigna/Ulrike Malmendier, ‘Paying Not to Go to the Gym’ (2006) 96 Am Econ Rev 694.

Consider what the study holds constant. Attendance is observable, controllable, and repeated: the consumer knows whether she went, chooses whether to go, and has months of her own data. Yet she overestimates her future consumption by a factor of two.

If she cannot predict whether she will go to the gym, what are the odds she can predict her token consumption?

Grubb and Osborne’s study of cellular service demand extends the finding.⁵² Consumers proved both inattentive to past usage and overconfident about future consumption. Critically, their model predicts that introducing bill-shock alerts – the intuitive remedy – would have *lowered* average annual consumer welfare by USD 33 once firms adjusted prices strategically in response.⁵³ Mandating disclosure without changing the underlying market structure may not merely fail to help; it may actively harm.

Recent Court of Justice jurisprudence has begun to acknowledge these limits. In *Compass Banca*, the Court held that the average consumer must still be defined ‘by reference to a consumer who is reasonably well-informed and reasonably observant and circumspect’, but that this definition ‘does not, however, exclude the possibility that an individual’s decision-making capacity may be impaired by constraints, such as cognitive biases’.⁵⁴

The ‘prudent decision’ standard of *D.V. v M.A.* must accordingly be read against the backdrop of bounded rationality. A consumer can be ‘reasonably observant and circumspect’ and still fail to process complex technical information. The question is not whether a hypothetically rational consumer *could* use disclosed information, but whether an actual consumer – subject to cognitive constraints the Court now recognises – *can*.

⁵² Michael D Grubb/Matthew Osborne, ‘Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock’ (2015) 105 Am Econ Rev 234.

⁵³ *ibid.* 234, 260–262.

⁵⁴ CJEU case C-646/22, *Compass Banca SpA v Autorità Garante della Concorrenza e del Mercato* ECLI:EU:C:2024:957, para. 59.

IV. The Variation Problem

Assume a (by now) parallel world in which both the measurement and prediction problems could be resolved: tokens are now a standardised measure, and a consumer possesses a reliable tool to forecast her future usage. She still faces a further obstacle: the limits themselves are not fixed, as virtually all providers reserve the right to adjust them during performance.

The Anthropic incident of July–August 2025 illustrates the problem concretely. On 28.7.2025, Anthropic announced via social media that it would introduce ‘new weekly rate limits’ for Claude Pro and Max subscribers, effective 28.8.2025, explaining that the limits targeted users ‘running Claude Code continuously in the background, 24/7’ and those engaging in ‘policy violations like account sharing and reselling access’.⁵⁵ The change was imposed during performance without any pre-contractual disclosure that such limits might be introduced.

Anthropic is not unusual. OpenAI’s release documentation acknowledges that it ‘periodically adjust[s]’ model parameters affecting usage, and that inferior models are substituted as a ‘fallback’ when rate limits are reached – without disclosing the limits themselves.⁵⁶ Microsoft’s Copilot terms reserve similar rights.⁵⁷

The pattern is uniform: every major provider retains unilateral authority to change what the consumer receives.

1. The Legal Framework for Usage Variation

The Court of Justice has addressed unilateral variation clauses in *Invitel* and *RWE Vertrieb*, establishing a two-limb test for compliance with the Unfair Contract Terms Directive.⁵⁸

⁵⁵ Anthropic (@AnthropicAI), post on X, 28.7.2025, <https://x.com/AnthropicAI/status/1949898502688903593> (accessed 10.9.2025).

⁵⁶ OpenAI, ‘ChatGPT Release Notes’, entries of 4.2.2026 and 18.3.2026, <https://help.openai.com/en/articles/6825453-chatgpt-release-notes> (accessed 7.4.2026).

⁵⁷ Microsoft, ‘Copilot Terms of Use’, <https://www.microsoft.com/en-us/microsoft-copilot/for-individuals/terms-of-use> (accessed 5.12.2025).

⁵⁸ CJEU case C-472/10, *Nemzeti Fogyasztóvédelmi Hatóság v Invitel Távközlési Zrt* ECLI:EU:C:2012:242; *RWE Vertrieb* (n 50) para. 49.

The first limb concerns foreseeability. The contract must set out transparently ‘the reason for and method of the variation’, so that ‘the consumer can foresee, on the basis of clear, intelligible criteria, the alterations that may be made’.⁵⁹ Pre-contractual disclosure is mandatory. As the Court explained, ‘[t]o the supplier’s legitimate interest in guarding against a change of circumstances there corresponds the consumer’s equally legitimate interest, first, in knowing and thus being able to foresee the consequences which such a change might in future have for him and, secondly, in having the data available in such a case to allow him to react most appropriately to his new situation’.⁶⁰ A term specifying no reason and no method gives the consumer nothing to foresee, though the ultimate unfairness assessment remains for the national court.⁶¹

The second limb concerns exit. The consumer must have an *effective* right of termination – one that is not merely formal but practicable given market conditions, switching costs, and notice periods. The national court must consider ‘whether the market concerned is competitive, the possible cost to the consumer of terminating the contract, the time between the notification and the coming into force of the new tariffs, the information provided at the time of that communication, and the cost to be borne and the time taken to change supplier’.⁶²

AI service variation clauses fail both limbs.

Regarding foreseeability: OpenAI’s statement that ‘limits may vary based on system conditions’ discloses neither the *reason* for variation (what conditions trigger adjustment?) nor the *method* (by what formula or within what range do limits change?). The consumer cannot foresee, on the basis of *any* criterion, what alterations may be made to her usage entitlement: ‘system conditions’ can mean, rather literally, anything.

EU law is no stranger to *acceptable* variation clauses. CPI indexation – the mechanism at issue in *AI Telekom Austria*, decided under Article 20(2) of the Universal Service Directive (rather than the UCTD) – was held not even to amount to a ‘modification to the contractual conditions’

⁵⁹ *RWE Vertrieb* (n 50) para. 49.

⁶⁰ *ibid.* para. 53.

⁶¹ *Invitel* (n 58) paras. 21–22 and 28.

⁶² *RWE Vertrieb* (n 50) para. 54.

triggering a termination right, because the index is ‘clear, comprehensive and easily accessible’ and ‘result[s] from State decisions and mechanisms’.⁶³ In other words, the law does recognise that *some* variation must be tolerated.

Three features explain why indexation survives where other variation clauses would fail. The reference is *objective* – measured by defined methodology, not provider discretion.⁶⁴ It is *external* – compiled by a public institution, not by the party benefiting from the adjustment. And it is *verifiable* – the consumer can confirm the published index value and calculate the resulting price change. (A fourth feature is somewhat implicit: the trader retains *no discretion* – the price moves mechanically with the index, and whether, by how much, and for whom is decided by no one.)

AI variation clauses fail on all the elements. ‘System conditions’ refers to no external standard. It is compiled by no public body. It is verified by no independent authority. And even if providers disclosed the triggers for adjustment, they would retain full discretion over the response – whether to throttle, by how much, for which subscribers, and for how long.

The pattern across regimes is consistent: the variation EU consumer law tolerates is mechanical, not discretionary. AI service variation is discretionary all the way down.

The reasoning in *Lexitor* – decided in the context of the Consumer Credit Directive rather than the UCTD, but applying an analogous transparency framework – takes the analysis, and the corresponding points, further.⁶⁵ In that case, the Court held that Article 10(2)(k) of Directive 2008/48 requires the conditions under which charges may be amended to be specified in a clear and concise manner, so that they ‘contain no imprecision objectively capable of misleading’ the consumer ‘as to the existence of events capable of triggering the change, and the link between

⁶³ Directive 2002/22/EC on universal service and users’ rights relating to electronic communications networks and services [2002] OJ L 108/51 (Universal Service Directive) Art. 20(2); *AI Telekom Austria* (n 25) paras. 26–29.

⁶⁴ See also Loos (n 14) 287 (‘a term should not leave the trader too much discretion in the exercise of his powers’, endorsing WH van Boom); Directive 93/13/EEC (n 11) Annex, points 2(c)–(d).

⁶⁵ CJEU case C-472/23, *Lexitor sp. z o.o.* ECLI:EU:C:2025:89, paras. 43–44 (interpreting Art. 10(2)(k) of Directive 2008/48/EC on credit agreements for consumers [2008] OJ L 133/66; applied here by analogy to the UCTD framework).

the change in charges and that event'. The Court identified as problematic conditions 'defined on the basis of indicators which are difficult for the consumer to verify, both before the conclusion of that agreement and during its performance', including 'the variable economic indicators' – among them those 'controlled by the bank itself'.⁶⁶

What went wrong for the terms in *Lexitor* was that the triggering indicators were unverifiable by the consumer and, in part, 'controlled by the bank itself'; the Court expressly held that caps on the amount and frequency of variation 'cannot invalidate that finding'. If a bounded variation governed by self-controlled, unverifiable triggers fails the test, an uncapped variation of the service allocation governed by 'system conditions' fails it *a fortiori*. The consumer cannot ascertain whether the triggering conditions have arisen, cannot calculate their effect on her allocation, and learns of the change only when she experiences it, defeating any semblance of an informed decision.

Rendering the remedies further ineffective in the context of AI services is the state of the market: switching to a competitor does not solve the problem when all competitors reserve identical variation rights. The competitiveness of the market that *RWE Vertrieb* identifies as relevant to effective termination *presupposes* meaningful differentiation.⁶⁷ Where all major providers maintain the same opaque variation clauses, the consumer's choice is not between transparent and opaque terms but between equivalent opacity at different providers.

The variation problem is, moreover, no longer governed solely by case law developed for other sectors. For a digital service supplied over a period of time, Article 19 of the Digital Content Directive regulates modification directly: the contract must allow the modification *and provide a valid reason* for it, and where the change negatively affects access or use by more than a minor degree, the consumer must be informed in advance, on a durable medium, of its features, its timing, and her right to terminate free of charge within 30 days.⁶⁸

⁶⁶ *ibid.* paras. 45–46.

⁶⁷ *RWE Vertrieb* (n 50) para. 54.

⁶⁸ Directive (EU) 2019/770 (n 10) Art. 19(1)–(2); by Art. 19(4), no termination right arises where the consumer may keep the unmodified service in conformity at no extra cost.

Yet Article 19 disciplines the *process* of announced change; it does not restore ex ante evaluability. It supplies no baseline against which a ‘negative impact’ could be quantified, it does not catch silent throttling that never surfaces as a datable modification, and it operates ex post, consumer by consumer – thus not solving the issue.

2. Structural Limits of Variation Disclosure

Even if providers disclosed their variation algorithms, the remedy would not cure the underlying defect. Here, once again, the structural features of AI services render variation mechanisms genuinely indeterminate – not merely undisclosed.

The *D.V. v M.A.* framework, discussed in Part III, establishes that inherent unpredictability does not relieve traders of transparency obligations – they must still provide information enabling prudent decisions.⁶⁹ And even the *RWE Vertrieb* framework does not require exact foresight. What it requires is a *method* – a disclosed mechanism linking variation to identifiable conditions. For CPI indexation, the method is a formula: price adjusts proportionally to the published index change. The consumer cannot predict the outcome; she can verify the mechanism. The question, therefore, is what an adequate method for AI service variation would look like.

True: providers could do better without doing perfectly. They could disclose historical variation patterns, specify ranges within which past adjustments have fallen, and commit to advance notification of changes affecting allocations. These measures would be imperfect but substantially better than ‘system conditions’. The absence of perfect foresight does not justify the absence of any foresight.

And yet: the deeper point is that disclosure alone *cannot* solve the variation problem – because the variation problem *compounds* the *measurement* problem: one cannot meaningfully disclose adjustment mechanisms for an allocation that was never quantified in the first place. If the consumer does not know what ‘5x more usage’ means, she certainly cannot evaluate what ‘adjusted system conditions’ will do to it. The opacity is hence layered: undefined baselines subject to undisclosed variation according to unspecified criteria.

⁶⁹ *D.V. v M.A.* (n 44) para. 43.

What can be done instead?

V. Comparative Label Transparency

Each failure documented so far has the same structure: disclosure hands the consumer information she cannot use – units she cannot compare, predictions she cannot make, mechanisms she cannot verify.

The transparency doctrine, as developed in *Kásler* and *D.V. v M.A.*, operates on a foundational assumption: that disclosed information enables the consumer to evaluate economic consequences.⁷⁰ The trader provides information; the consumer processes it; a prudent decision follows. The framework accommodates complexity and uncertainty, as Part II showed. What it assumes, in both cases, is that the information, once provided, is *usable*. The consumer may need effort or assistance to use it, but the information bears a cognisable relationship to the decision she must make.

1. 1. The Many Failures of Disclosure

For AI services, this assumption fails – not simply because the information is too complex, but because the consumer cannot verify the amount and appropriateness of the computational resources used. This is the defining characteristic of what economists term credence goods: goods whose qualities remain unverifiable even after consumption.⁷¹ The classic example is the expert service – the mechanic who replaces an engine block when a minor repair would suffice.⁷² The consumer receives the service; she cannot assess whether the service was proportionate to her need.

The verification failure, when it comes to AI services, operates on two independent levels. At the first level – meter accuracy – the consumer cannot audit whether consumption was *measured correctly*. If a provider states ‘you used 80% of your allocation today’, the consumer

⁷⁰ Mateja Đurović, ‘How to Remain Fit for the Digital Age: The Example of EU Consumer Law’ (2024) 13 EuCML 249, 251 (digitalisation has shifted the problem from information scarcity to surplus); the third failure identified in the text – indeterminacy – is this Article’s own extension.

⁷¹ Michael R Darby/Edi Karni, ‘Free Competition and the Optimal Amount of Fraud’ (1973) 16 J Law & Econ 67, 68–69.

⁷² *ibid.* 69.

cannot confirm that each query was metered accurately, nor that system overhead was appropriately allocated rather than inflated. There is no external meter; the provider both supplies the service and operates the only measurement instrument. At the second level – consumption necessity – even if the meter were perfectly honest, the consumer could not assess whether the consumption was *proportionate*. Could the system have answered her query with fewer tokens? The consumer cannot know – and, critically, she cannot know even *after* receiving the service. This is not a gap that retrospective disclosure could fill. The meter problem and the necessity problem are analytically distinct, but both produce the same regulatory consequence: the consumer’s inability to verify what she has received renders conventional disclosure ineffective.

The distinction from the cognitive limitations discussed in Part III is clear, and consequential for regulatory design. Cognitive biases – overconfidence, inattention, misprediction – are failures of consumer processing, and in principle, a perfectly rational consumer could overcome them. The verification gap is different: the information required to assess whether consumption was appropriate does not exist in any form accessible to the consumer – a typical credence good – as it resides inside provider systems, measured by provider methods, reported at provider discretion.

This distinction must be taken into account when discussing transparency.

Where the problem is cognitive, disclosure reforms may help: simplification, standardisation, salient presentation. Where the problem is epistemic, disclosure reforms cannot help, however well designed. What is needed instead is institutional substitution:⁷³ replacing consumer verification with external certification. Experimental research by Dulleck, Kerschbamer and Sutter confirms the regulatory implication: in credence markets, verifiability alone – enabling consumers to observe what service was performed – had ‘at best a minor ... effect’ on market outcomes, while a structural intervention – liability – significantly improved them.⁷⁴ Building

⁷³ Christine Riefa, ‘Protecting Vulnerable Consumers in the Digital Single Market’ (2022) 33 EBLR 607, 607–608 (arguing the information paradigm should be reversed).

⁷⁴ Uwe Dulleck/Rudolf Kerschbamer/Matthias Sutter, ‘The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition’ (2011) 101 Am Econ Rev 526, 531, 538.

verification infrastructure is insufficient. The remedy must *replace* consumer evaluation, not supplement it.

This aligns with broader scholarship on mandated disclosure. Ben-Shahar and Schneider’s systematic critique concludes that mandated disclosure ‘chronically fails to accomplish its purpose’ and produces ‘unintended and undesirable consequences, like driving out better regulation and hurting the people it purports to help’.⁷⁵ The failures, they argue, are not random but structural.

The failure of disclosure does not, however, exhaust the transparency doctrine. It redirects inquiry towards mechanisms that do not depend on the consumer’s comprehension of technical detail – or on her capacity to verify the usage.

2. The Comparative Transparency Model

Fortunately, the EU has addressed analogous problems elsewhere – if not at a single obvious place.

The pattern it follows is consistent: where consumers cannot evaluate quality or quantity directly, institutional infrastructure provides comparative evaluation capacity without requiring technical understanding. The European Electronic Communications Code mandates standardised contract summaries enabling cross-provider comparison of telecommunications services;⁷⁶ the PRIIPs Regulation translates packaged retail investment products into Key Information Documents with standardised risk indicators.⁷⁷ The extension to AI services follows the same logic.

The EU Energy Label is the clearest precedent. Regulation 2017/1369 establishes a framework for labelling energy-related products – refrigerators, washing machines, air conditioners –

⁷⁵ Omri Ben-Shahar/Carl E Schneider, ‘The Failure of Mandated Disclosure’ (2011) 159 U Pa L Rev 647, 651.

⁷⁶ Directive (EU) 2018/1972 establishing the European Electronic Communications Code [2018] OJ L 321/36, Art. 102(3).

⁷⁷ Regulation (EU) No 1286/2014 on key information documents for packaged retail and insurance-based investment products [2014] OJ L 352/1.

according to standardised efficiency classes.⁷⁸ The label displays a simple A-to-G scale; consumers need not understand compressor technology or thermodynamic efficiency to compare products.

The system works through institutional substitution – replacing consumer evaluation with institutional certification – and offers clear benefits over the mandated single disclosure.

First, standardised measurement serves as a baseline. The Regulation requires that energy consumption be measured according to harmonised European standards, using specified test conditions and calculation methodologies.⁷⁹ A kilowatt-hour measured by one manufacturer is commensurable with a kilowatt-hour measured by another. Different products with different internal architectures receive comparable ratings because the benchmark measures outcomes, not mechanisms. The unit is standardised across the market – which is precisely what AI tokens lack.

Second, independent verification solves the credence good issue. Manufacturers self-declare energy efficiency, but Member States conduct market surveillance testing, and products exceeding declared consumption beyond permitted verification tolerances face enforcement action;⁸⁰ a public registry allows marketed claims to be checked.⁸¹ The system does not rely on consumer verification; public authorities verify on consumers' behalf.

Third, comparative presentation allows for informed decisions. The label translates technical measurements into decision-relevant comparisons. A consumer comparing two refrigerators need not calculate annual energy costs; she sees that one is rated B and the other D. The rating absorbs technical complexity and delivers comparative evaluation capacity.

The transparency achieved is, strictly speaking, not informational – but *comparative*. The consumer does not understand *how* the products differ technically, but understands *that* they differ on a standardised, verified scale.

⁷⁸ Regulation (EU) 2017/1369 setting a framework for energy labelling [2017] OJ L 198/1.

⁷⁹ Commission Delegated Regulation (EU) 2019/2014 [2019] OJ L 315/29, Annex IV; EN 60456 per Commission Implementing Decision (EU) 2021/936 [2021] OJ L 204/42, Annex I.

⁸⁰ Regulation (EU) 2017/1369 (n 78) Arts. 3(3), 8.

⁸¹ *ibid.* Arts. 4 and 12.

The system's effectiveness is well-documented. In a 2024 Eurobarometer survey, 75% of respondents reported that, in the preceding five years, the label had influenced their choice when purchasing an appliance.⁸² Schleich, Durand, and Brugger's analysis of sales data across eight EU countries found that the combined effect of the 2010–2011 minimum energy performance standards and accompanying label changes increased the sales share of efficient appliances by 15–38 percentage points – a population-level behavioural shift vastly exceeding effects achievable through technical energy disclosure alone.⁸³

3. Application to AI Services

The solution thus seems simple: a comparative transparency regime for AI services would require the same three structural elements – standardised measurement, independent verification, and comparative presentation – but the application must be tailored to the specific failures the preceding Sections identified.

Output *quality* – whether responses are accurate, helpful, or well-reasoned – is an experience-good dimension beyond the scope of this Article. Consumers can evaluate through use: every user will form a view on whether the model's writing is good, whether its coding suggestions work, whether its analysis is insightful. These assessments are subjective and task-dependent; no label should rate them.

Still, when it comes to usage capacity, the Energy Label model could be readily translated to AI services. An independent technical body – drawing on standardisation organisations such as CEN or CENELEC – would design task-based benchmarks reflecting typical consumer use cases: research queries, coding assistance, document analysis, creative writing. For each task category, the benchmark would measure how many representative tasks each provider's subscription tier permits before exhausting a typical monthly allocation. The measurement absorbs tokenisation differences, reasoning-token overhead, and context-window effects. The

⁸² European Commission, 'Special Eurobarometer 555: Europeans' Attitudes towards EU Energy Policy' (2024) 60–61 (QC11).

⁸³ Joachim Schleich/Antoine Durand/Heike Brugger, 'How effective are EU minimum energy performance standards and energy labels for cold appliances?' (2021) 149 Energy Policy 112069 (combined effect of MEPS and label changes; contributions not separable).

measurement problem – the absence of a common unit enabling cross-provider comparison – is dissolved, not by standardising tokens but by rendering them irrelevant.

The benchmark would also address the verification gap identified in Section V.1. If a provider's system generates verbose responses consuming excessive tokens while a competitor achieves equivalent results concisely, the benchmark exposes this. Systematic overtreatment – the credence-market failure Darby and Karni identified – would produce poor efficiency ratings. The consumer cannot audit her own consumption; the certifying body can. And – critically – it can also audit the meter, verifying that reported consumption corresponds to actual processing rather than inflated overheads.

Benchmark results would be translated into simplified categories – an A-to-G scale or equivalent. The ratings would be task-specific, because value varies by use case: a tier optimised for coding assistance may deliver less value for extended research conversations. A consumer comparing subscriptions would see: 'Provider A's EUR 20/month tier is rated B for research tasks, C for coding tasks. Provider B's EUR 25/month tier is rated A for research, B for coding.' She understands that B is better than C and worse than A. The prediction problem – the consumer's inability to forecast consumption – is displaced rather than solved: for the benchmarked task categories, the rating states what the subscription permits, removing the need to forecast what it will cost.

A certifying body could verify that marketed tiers deliver the rated performance. Verification would operate on two dimensions: *ex ante*, the body would test subscription tiers against benchmarks before certification; *ex post*, it would conduct ongoing market surveillance – purchasing subscriptions, running test queries, confirming that delivered performance matches rated performance. Significant deviations would trigger rating downgrades, public notices, or – for systematic misrepresentation – removal from the certification scheme. The Benchmarks Regulation's governance model – methodology transparency, conflict management, audit trails – transfers well to AI service benchmark administration.⁸⁴

⁸⁴ Regulation (EU) 2016/1011 on indices used as benchmarks in financial instruments and financial contracts or to measure the performance of investment funds [2016] OJ L 171/1.

Naturally enough, certification cannot be static. AI models are updated frequently; capacity allocations shift; a rating that accurately captured performance in January may mislead by March – a genuine disanalogy with the Energy Label’s static ratings, but not a disabling one. EURIBOR is recalculated daily; Euro NCAP may retest modified vehicles under specified conditions.⁸⁵ The principle that certification must track changing performance is not novel; it requires institutional capacity, not conceptual innovation. Providers would bear ongoing obligations to notify the certifying body of changes affecting rated performance and to re-run benchmark tests upon material modification. Rating changes would be published, ensuring that variation – the problem identified in Part IV – becomes visible through an institutionally verified channel rather than hidden behind ‘system conditions’. The lag between performance change and rating update is a limitation – but it is a limitation of degree, not of kind. A rating system that captures performance with a one-month delay is imperfect; a market with no comparative information at all is opaque.

Doctrinally, the rating system does not itself supply the reason for and method of variation that *RWE Vertrieb* requires the contract to set out – no external metric can discharge an obligation that attaches to the drafting of the term, and the judgment forecloses substituting in-performance notice and exit for pre-contractual disclosure. What the rating supplies is the precondition for discharging that obligation. A provider can commit, in the contract, that the service will maintain its certified rating and that any modification affecting it will be reflected in the published rating, notified in advance, and accompanied by a penalty-free termination right – converting service degradation from an unverifiable private experience into a datable, public event, and giving the required disclosure of reason and method something to bite on.

That construction serves the two consumer interests the Court identifies even where the provider’s internal triggers resist formulaic ex ante statement. And it rehabilitates the second limb directly: paragraph 54 makes termination effective only where the consumer can verify the change and the market offers a genuine alternative, and both conditions presuppose exactly the

⁸⁵ European Money Markets Institute, ‘Euribor’, <https://www.emmi-benchmarks.eu/benchmarks/euribor/> (accessed 22.7.2026) (rates calculated and published for each fixing day); Euro NCAP, ‘Vehicle Specification, Sponsorship, Testing and Retesting Protocol’ (v7.4.3, 2021) para. 6.2.1.2, <https://www.euroncap.com/protocols/> (accessed 22.7.2026).

comparative infrastructure the rating provides – curing the equivalent-opacity failure identified in Part IV.⁸⁶ The contractual statement of a valid reason remains required, by point 1(j) of the Annex to the UCTD and Article 19 DCD alike; the rating supplements that requirement rather than supplying it, but it transforms compliance from an unverifiable assertion into a commitment whose observance the consumer can check.

Again, the displacement is partial: the consumer deviating from benchmarked workflows still cannot know how quickly she will exhaust her allocation. Usage alerts, modelled on the Roaming Regulation, would address this gap during performance: Article 14(4) of Regulation 2022/612 requires operators to notify customers at 80% of the agreed limit and again before it is exceeded.⁸⁷ The consumer who receives an 80% alert can moderate usage, prioritise remaining capacity, or investigate upgrades.

This does not cure the prediction failure – but it provides the ‘periodic reports’ that *D.V. v M.A.* identified as an alternative to ex ante estimation.⁸⁸

The AI Act – though, again, not applicable in substance – provides a pathway which could inspire implementation. Article 95 enables the AI Office and Member States to encourage voluntary codes of conduct for AI systems other than high-risk AI systems (Article 95(1)), including codes drawn up on the basis of ‘clear objectives and key performance indicators to measure the achievement of those objectives’ (Article 95(2)).⁸⁹ Consumer subscription services fall outside the high-risk classification, and Article 95 does not contemplate consumer task-capacity labels as such. What it supplies is a docking point rather than a mandate: an expressly non-exhaustive framework within which the AI Office and Member States must encourage voluntary codes for non-high-risk systems, developed with the involvement of consumer protection organisations – onto which a benchmarking code of the kind proposed here could readily be grafted.

⁸⁶ *RWE Vertrieb* (n 50) paras. 51, 53–54.

⁸⁷ Regulation (EU) 2022/612 on roaming on public mobile communications networks within the Union [2022] OJ L 115/1, Art. 14(4).

⁸⁸ *D.V. v M.A.* (n 44) para. 44.

⁸⁹ Art. 95(1)–(3) AI Act; recital 165.

Nor does Article 95 stand alone: the Act repeatedly presupposes the measurement infrastructure such a code would require. Article 15(2) directs the Commission, in cooperation with ‘metrology and benchmarking authorities’, to encourage ‘the development of benchmarks and measurement methodologies’; and, at the model layer on which consumer subscriptions run, capability benchmarking is already an operative legal technique – systemic-risk classification proceeds by reference to ‘benchmarks and evaluations of capabilities of the model’, and the AI Office is empowered to conduct model evaluations through API access.⁹⁰

Article 112(7)’s triennial review of the codes of conduct supplies the pathway through which a voluntary label, once proven, could harden into binding form.⁹¹

The institutional infrastructure, moreover, is not hypothetical: CEN-CENELEC Joint Technical Committee 21 is already developing harmonised AI standards under the Commission’s standardisation request,⁹² and ANEC, the European consumer voice in standardisation, sits in all five working groups.⁹³

In other words, the capacity exists.

VI. Conclusion

Three structural features of AI services defeat the conventional transparency approach of EU consumer protection law. As shown, consumers cannot reliably measure their usage, predict their consumption, or detect variation in the service itself – nor verify, even retrospectively, that the resources consumed corresponded to the tasks performed. In each case, the deficit is only partly informational – and, as Parts II and IV have shown, only partly curable by the remedial layer of the *acquis*. The DCD’s conformity and modification rules, and the CRD’s integration

⁹⁰ Arts. 15(2), 51(1)(a), 66(g), 68(3)(a)(ii) and 92 AI Act; Annex XIII, point (e). Art. 15(2) is invoked as infrastructure a voluntary code could borrow, not as a direct legal basis.

⁹¹ Art. 112(7) AI Act.

⁹² Commission Implementing Decision C(2023) 3215 final of 22.5.2023; repealed and replaced by Commission Implementing Decision C(2025) 3871 final of 23.6.2025 (high-risk AI systems).

⁹³ ANEC, ‘Digital Society’, <https://anec.eu/priority/digital-society/> (accessed 7.4.2026); on the AI Taskforce, CEN-CENELEC, ‘ANEC launches AI Taskforce’, On the Spot 62 (1.5.2025), <https://www.cencenelec.eu/news-events/news/2025/newsletter/ots-62-anec/> (accessed 19.7.2026).

rule, sanction what providers state and how they may change it; but their triggers presuppose a measurable baseline that does not exist, and they operate ex post, dispute by dispute.

The residue is at its core architectural: more or better disclosure cannot help where the information needed for consumer evaluation either does not exist in accessible form or cannot be processed even if provided.

The comparative transparency model reframes the legal analysis.

Current doctrine asks whether the *consumer* can evaluate economic consequences. The Energy Label model asks whether the *system* enables evaluation – whether institutional infrastructure translates provider characteristics into consumer-accessible comparison. This reframing finds support in the Court’s jurisprudence. *Justa* held that transparency is satisfied when consumers can evaluate economic consequences; it did not require that they achieve this evaluation through their own unaided comprehension. *AI Telekom Austria* treated price variation tied to an objective public index as raising no modification concern at all: an external, institutionally verified mechanism can answer foreseeability concerns. *D.V. v M.A.* required information enabling a ‘prudent decision’ – and a prudent decision does not demand technical mastery; it demands a sufficient basis for evaluation.⁹⁴

Comparative benchmarks provide that basis. They would also supply what the existing remedial layer lacks: a standard for what is ‘normal’ for digital services of the same type, and a yardstick for when a modification’s impact is more than minor. The proposal does not compete with the conformity and modification rules already in force; it is the missing precondition for their operation. Where product complexity defeats direct comprehension, institutional infrastructure can enable evaluation indirectly. The question is whether the system – not the individual – provides the means.

The Article opened with an electricity tariff no consumer would sign. The remedy, fittingly, is the instrument European law built for electricity: not a fuller contract, but a label.

⁹⁴ *D.V. v M.A.* (n 44) para. 43.