Five ways leaders can make AI token usage sustainable at scale

As AI scales, token usage becomes more than a technology metric. It directly influences cost, operational performance and, increasingly, energy demand and emissions. A sustainable approach is not about reducing usage. It is about managing how AI scales so financial, operational and environmental outcomes move together.

1. Treat token usage as a shared operating input

Token consumption is not owned by a single function. It acts as a metered input that simultaneously affects:

- Financial performance (cost and margin)
- System behavior (efficiency and design)
- Environmental impact (energy use and emissions)

At scale, this shifts AI from a technical capability to a shared operating constraint. Decisions about how AI is used now have consequences across finance, technology and sustainability, not just within one domain.

---

2. Plan for variability across cost, performance and impact

As AI usage grows, token consumption introduces variability into:

- Operating costs
- System performance
- Energy demand

These variables move together, but not always predictably. A change that improves performance may increase cost. A reduction in cost may shift work back to people. Lower consumption does not automatically translate into lower environmental impact.

Leaders are not just managing spend. They are managing the interdependence of cost, output quality and resource demand at scale.

3. Focus on value, not just consumption

Token usage only creates value when it produces useful outcomes. When it does not, it creates multiple forms of waste:

- Financial waste (token spend with no return)
- Operational waste (rework and retries)
- Environmental waste (energy consumed without productive output)

At scale, these inefficiencies compound. Reducing visible consumption alone does not solve the problem if value per interaction does not improve. Sustainable token usage depends on aligning consumption with meaningful outcomes.

4. Make token usage visible across all three dimensions

Visibility is not just about cost tracking. It is about understanding how token usage behaves across:

- Financial performance
- System efficiency
- Environmental impact

Without a shared view, each function sees only part of the problem:

- Finance sees volatility in spend
- Technology sees system inefficiency
- Sustainability sees rising energy use

At scale, fragmented visibility creates conflicting decisions. A sustainable approach depends on integrated visibility that connects usage to cost, performance and impact simultaneously.

5. Align decision-making across finance, technology and sustainability

Token sustainability ultimately depends on how trade-offs are managed.

- Finance prioritizes predictability and margin
- Technology prioritizes performance and system design
- Sustainability prioritizes energy use and emissions

At scale, these priorities intersect in every AI-driven workflow. Without alignment, decisions can optimize for one outcome while weakening another. With alignment, organizations can scale AI in a way that is financially viable, operationally reliable and environmentally responsible.

Contacts

Joe Depa
EY Global Chief Innovation Officer

Colm Devine
EY Global Vice Chair – Sustainability

Pablo Cebro
EY Global Head of Technology Platforms, Client Technology

EY | Building a better working world

EY is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fueled by sector insights, a globally connected, multidisciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

All in to shape the future with confidence.

BMC Agency

GA 15516563

EYG no. 113444-26-Gbl

ED None

ey.com