

In the rapidly evolving world of AI-assisted products, one question repeatedly emerges from product managers, designers, and engineers alike: *How can we give users control over AI? How do we let them dial it up or down?* This isn't just about fancy sliders or toggles. It's about creating user control that respects individual preferences and task contexts while balancing automation's potential with human judgment.

Companies like **PM Toolkit**, **PayPal**, and **H2O.ai** have approached these challenges thoughtfully, providing valuable lessons for anyone shipping AI-powered features. Leveraging tools like *PM Toolkit calculators* for robust evaluation and spreadsheets to prototype lightweight workflows is essential to avoid common pitfalls.



The Model Is Not the Product

Before we dive into implementing user controls for AI behavior, let's clear a fundamental misconception often heard in AI product discussions:

The model is not the product.

AI models are powerful components but not the whole story. The product comprises the entire user journey, workflow, interface, and contextual safeguards around AI output. For example, **PayPal** uses AI extensively for fraud detection but layers the model's output with user alerts, risk scores, and human review options. It's not just about "accuracy" or "speed" — it's about trust, control, and reliability in the product.

As a product leader turned AI product lead, I always ask: **What do users do today?** How are they currently solving their problems? This guides us from vague "model wars" into concrete, task-specific AI behavior adjustments.

Why Give Users a Dial?

Giving users a way to tune AI behavior—like a dial to turn AI "up" or "down"—is rooted in three key themes:

- **User Control:** Empower users to feel in charge of AI impact, reducing friction and building trust.
- **Task-Dependent Behavior:** Different tasks need different AI modes and failure tradeoffs.

- **Transparency & Opt-Out:** Let users easily opt out or adjust settings in response to mistakes or unexpected outputs.

Let's break these down with examples and frameworks.

Define an Evaluation Before Release

Teams at **H2O.ai** and oft-discussed PM best practice hub **PM Toolkit** stress that you cannot ship AI features effectively without a defined evaluation plan.

This means specifying, up front, the following:

1. **What success looks like:** Define concrete goals using metrics relevant to your users and business. Use PM Toolkit calculators such as *TAM/SAM/SOM* to size your market, *ROI and business case calculators* to justify investment, and *RICE or ICE scoring* to prioritize features.
2. **Task-specific behavior:** Develop lightweight evaluation spreadsheets to test AI outputs in realistic contexts. Assess failure modes, accuracy thresholds, and user error tolerance.
3. **Feedback loops and kill switches:** Always build a kill switch and quick rollback paths. Set up A/B tests with sample size calculators from PM Toolkit to verify hypotheses before full rollout.

For example, if you build an AI autocomplete for customer support replies, define what error rate users tolerate before they get annoyed and switch the AI off. Your dial should give them control to adjust for that.

Task-Dependent Model Behavior & Failure Modes

AI models behave differently depending on the task and data context. Some workflows benefit from aggressive automation (like classification), others from augmentation (like-assisted writing), and some represent genuine innovation (like personalized recommendations).

Understanding these patterns is critical in defining your "dial" controls. Here are three archetypes:

Pattern	Description	User Control	Use Case	Failure Mode	Considerations
Automation	AI takes full action with minimal user involvement. Users can toggle AI to act autonomously or manually approve decisions. False positives can cause costly errors; include kill switches and alert systems.				
Augmentation	AI assists users by providing suggestions or insights. Users adjust AI "frequency" or suggestion aggressiveness. Over-reliance on suggestions may cause complacency; allow opt-out.				
Innovation	AI adds novel capabilities previously impossible. Users choose between traditional and AI modes, or blend them. Unpredictable behavior may confuse users; transparency and onboarding help.				

At **PayPal**, AI fraud detection uses automation for flagging transactions but empowers users with override or dispute options—an excellent example of balancing automation with user control.

Giving Users the Dial: Practical Approaches

How do you translate these concepts into actual product features? Here are several approaches that **PM Toolkit** users and AI product teams at **H2O.ai** and others iterate on:

1. Multi-Level Settings

Offer multiple settings layered by expertise and risk tolerance. For example:

- **Basic:** Set AI auto-suggestions to low frequency or passive mode.
- **Intermediate:** Allow medium frequency with user review required.
- **Advanced:** AI can auto-act with minimal intervention.

This caters to diverse user archetypes, from cautious novices to power users.

2. Contextual Dials Per Task

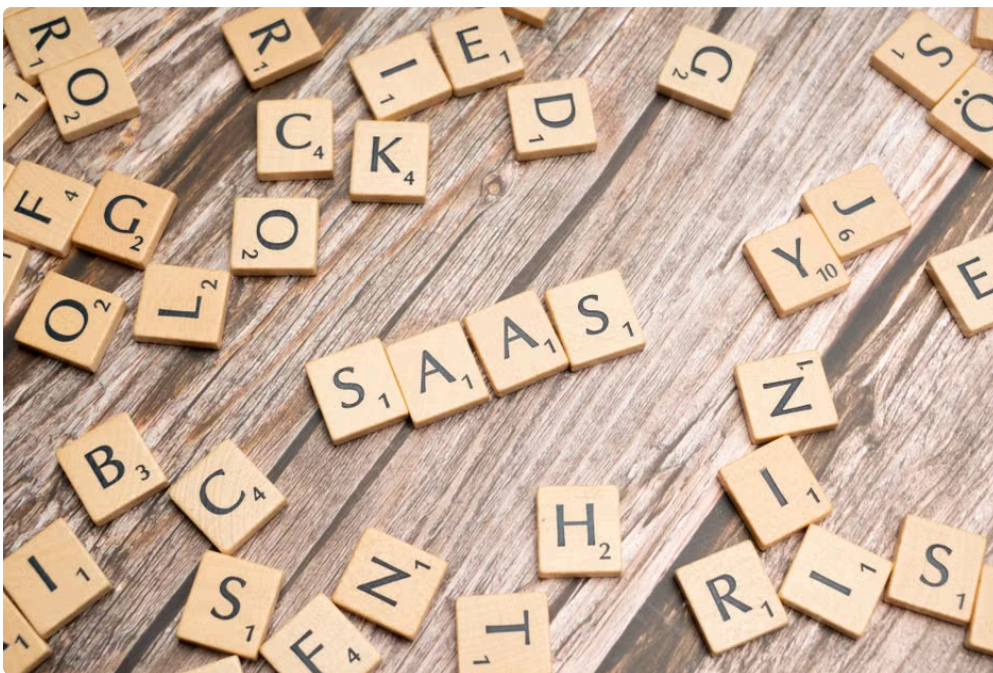
AI impact varies per feature. Instead of one universal slider, create task-specific controls. For instance, a marketing tool can have separate dials for:

- AI-generated email subject lines
- Content tone and creativity level
- Frequency of AI-driven campaign suggestions

This granularity gives users nuanced control and reduces cognitive overload.

3. Clear Opt-Out Options

Always include an easy way to disable AI features entirely. This respects users' preference for manual control and handles trust issues or poor AI performance gracefully.



4. Real-Time Feedback and Undo

Couple the dial with immediate previews or confidence scores. Allow users to revert AI actions swiftly if outcomes don't meet expectations.

5. Explainability and Transparency

Integrate inline explanations on what turning the dial means. For example:

"Increasing AI automation will apply suggestions automatically but may miss context in rare cases." <https://pmtoolkit.ai/learn/ai-modern-pm/ai-literacy-for-pms>

Such transparency helps users make informed choices.

Evaluating Your AI Dial Feature

Avoid launching without quantifiable data. Use **PM Toolkit calculators** to plan and assess early results:

- **Sample Size:** Calculate how many users need to test to reach statistical significance.
- **A/B Testing:** Run experiments comparing different dial defaults or presence vs absence.
- **NPS & PMF Scores:** Measure user satisfaction and product-market fit impact from control features.
- **Retention/Churn:** Evaluate if giving control affects long-term engagement.
- **Velocity and Cycle Time:** Assess if AI settings influence user efficiency positively or negatively.

Spreadsheets remain invaluable for lightweight evals during early stages, enabling you to iterate on failure modes and dial tuning quickly without complex ops overhead.

Conclusion: Design With Empathy and Data

Giving users the ability to turn AI “up” or “down” is not a mere UI gimmick—it’s a thoughtful product decision that requires understanding users, their workflows, and trust dynamics. Remember:

- The AI model is a component, not the whole product.
- Define concrete evaluation criteria before shipping.
- Customize controls per task and user sophistication.
- Prioritize transparency, feedback, and opt-out mechanisms.
- Use proven tools like **PM Toolkit calculators** and spreadsheets for iterative evaluation.

By combining these principles, inspired by lessons from innovators like **PayPal** and **H2O.ai**, you empower your users—not just your AI—to deliver meaningful value. After all, AI should augment human intelligence, not replace human agency.

What “dial” controls have you experimented with in your AI products? Share your stories and challenges!