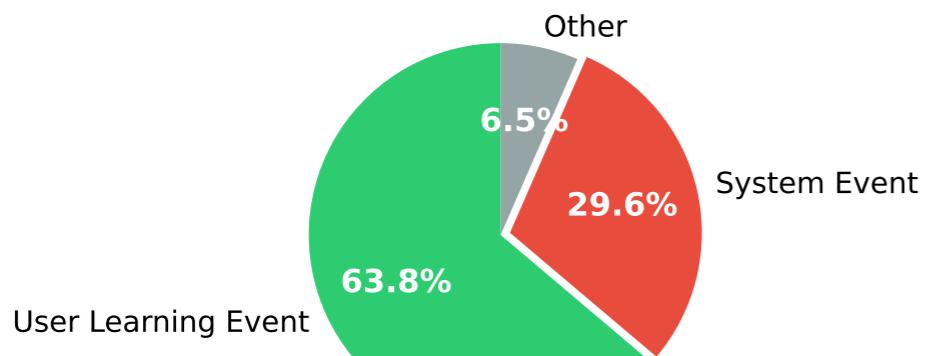


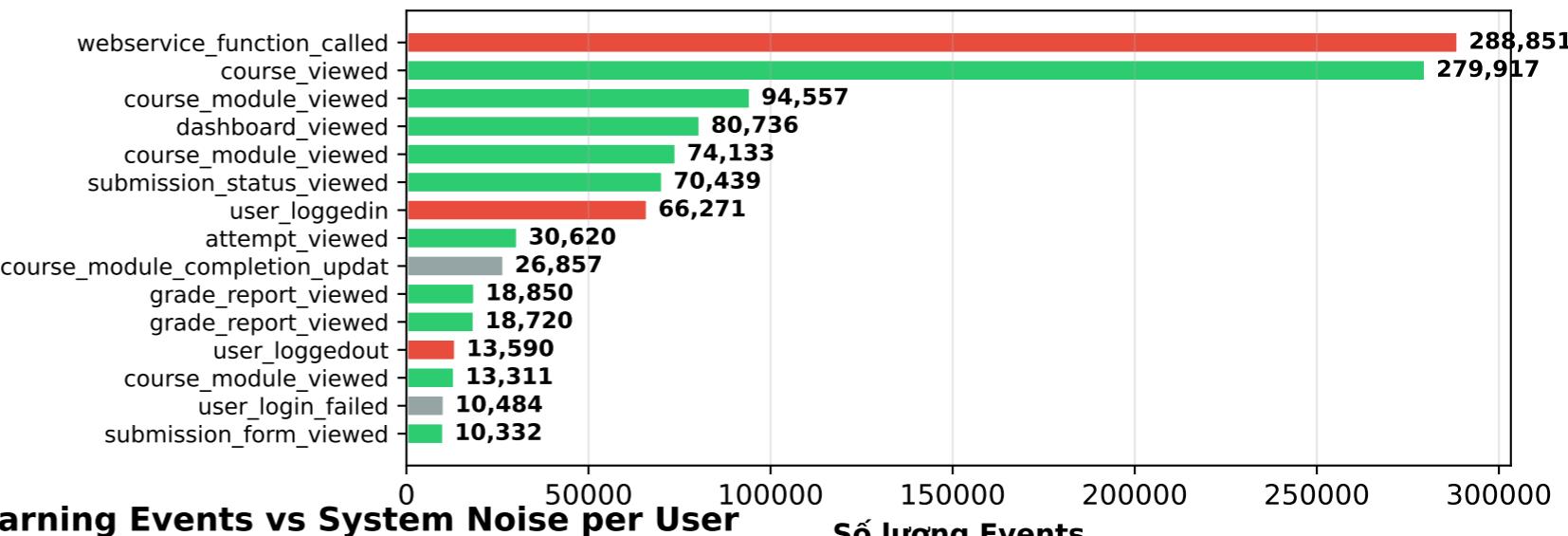
# KHÓ KHĂN 1: SYSTEM NOISE (Nhiều Hệ thống)

## Dữ liệu chứa nhiều sự kiện tự động không phản ánh hành vi học tập thực

### Tỷ lệ System Noise trong Dataset

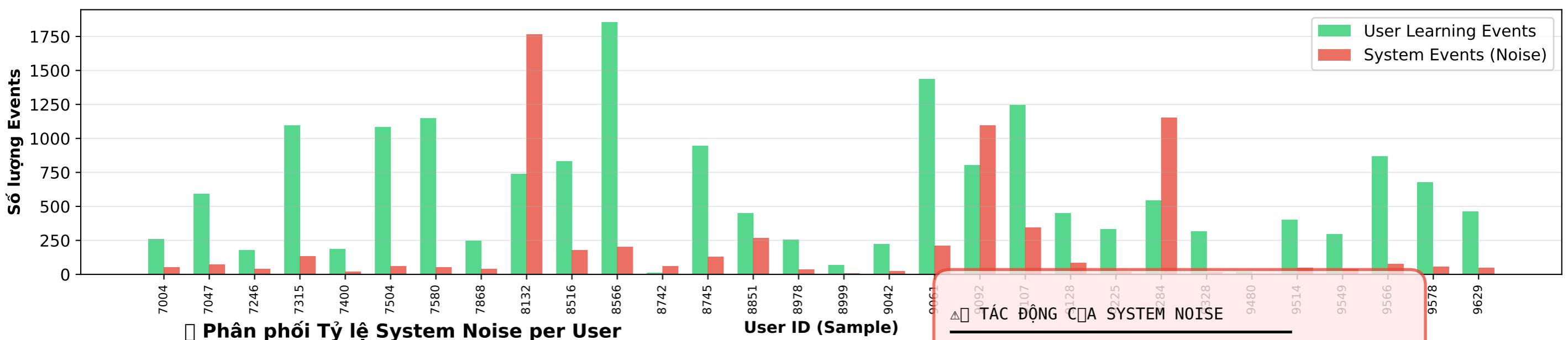


### Top 15 Event Types (Đỏ = System Noise, Xanh = User Learning)

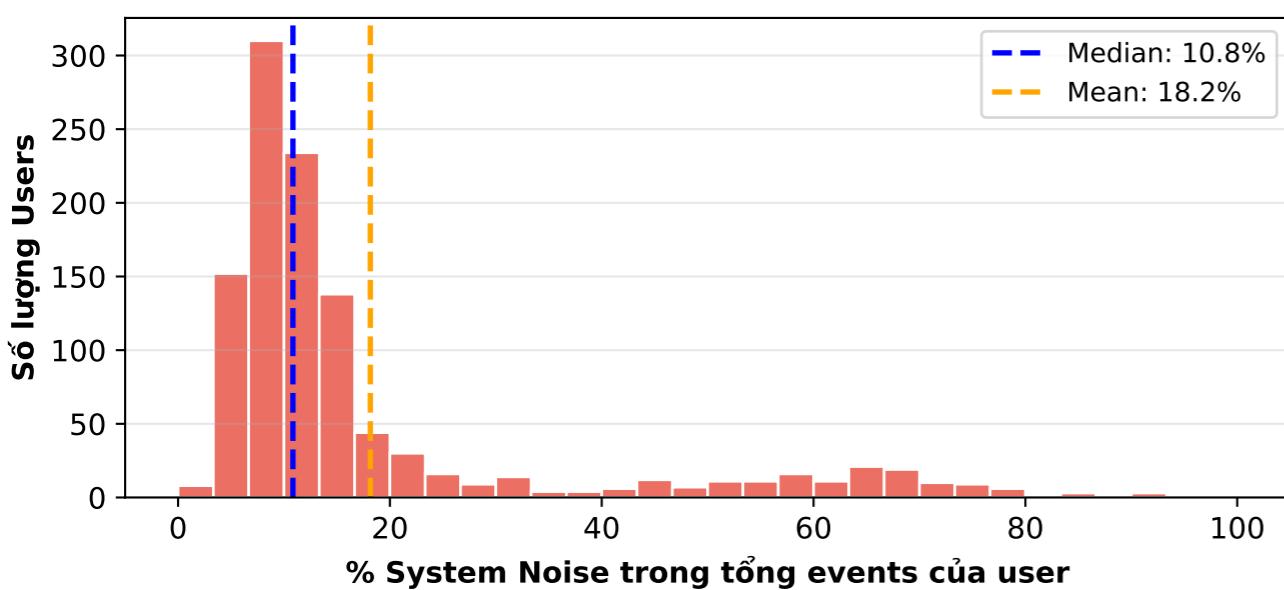


⚠ So sánh: Learning Events vs System Noise per User

(System noise có thể vượt quá learning events thực tế!)



### Phân phối Tỷ lệ System Noise per User



Median: 10.8%  
Mean: 18.2%

Thống kê:  

- System Events: 373,173 (29.6%)
- Learning Events: 803,770 (63.8%)
- Tỷ lệ noise trung bình/user: 18.2%

Vấn đề:  

- Làm sai lệch chỉ số "hoạt động" của học viên
- User có nhiều system events ≠ học chăm chỉ
- Khó hướng clustering: gom nhóm nhầm
- Khó hướng recommendation: đứt xuôi sai

Giai pháp:  

- Lọc bỏ system events trước khi phân tích
- Chỉ dùng learning-related events
- Tạo feature "learning\_intensity" riêng
- Weighted events (learning events > system)