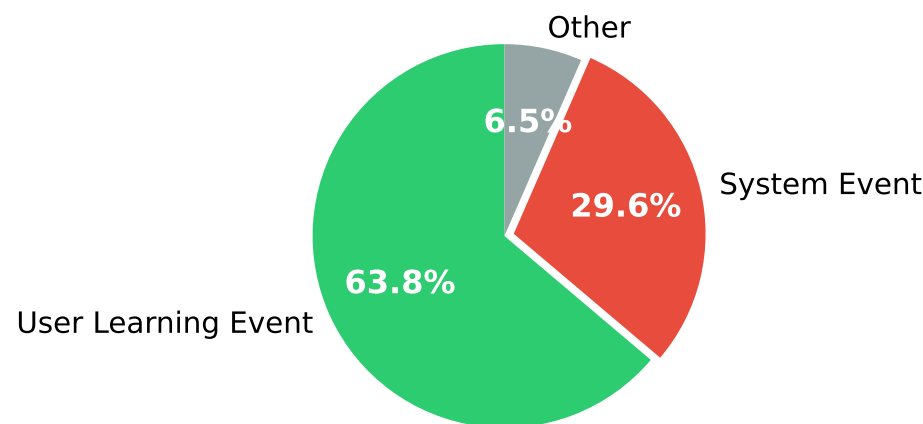


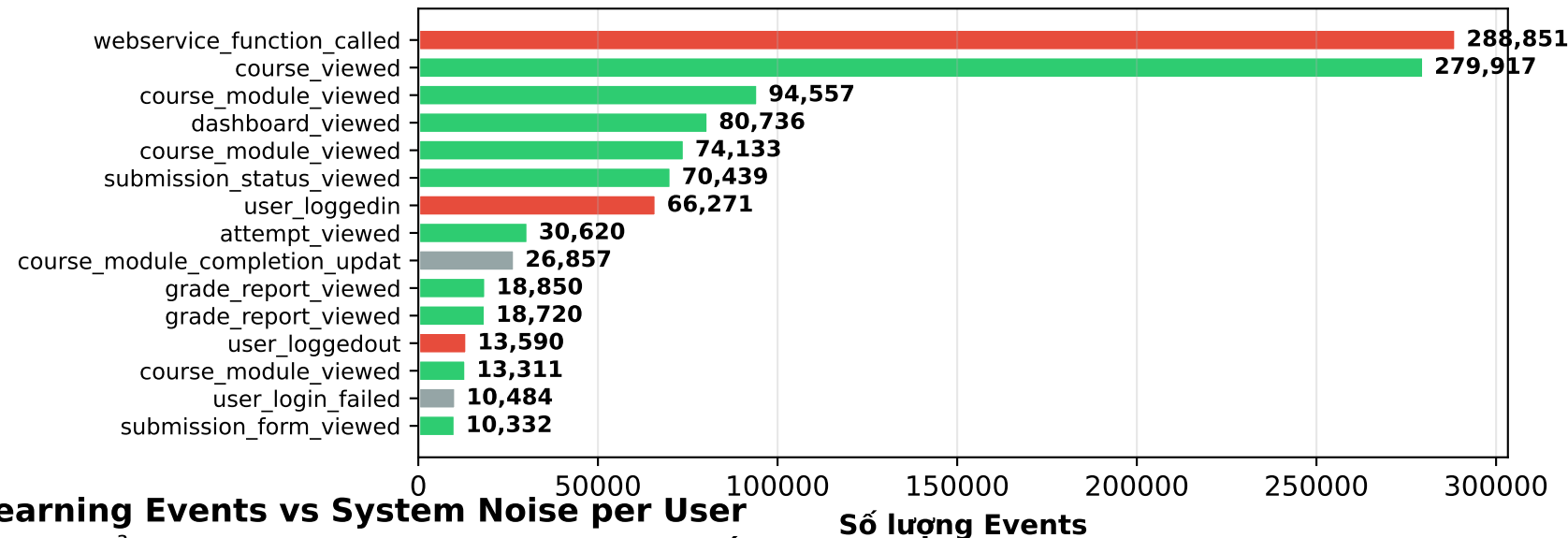
# 🔴 KHÓ KHĂN 1: SYSTEM NOISE (Nhiều Hệ thống)

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

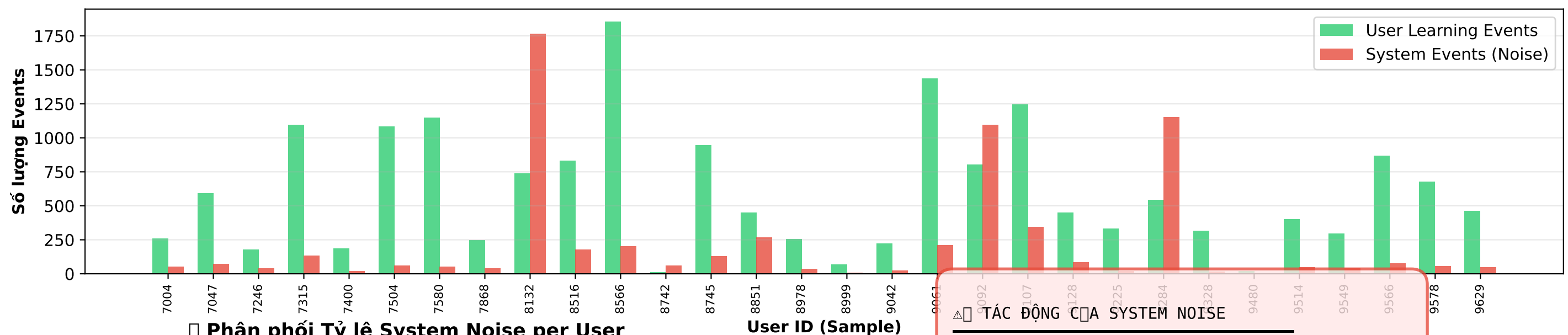
📊 Tỷ lệ System Noise trong Dataset



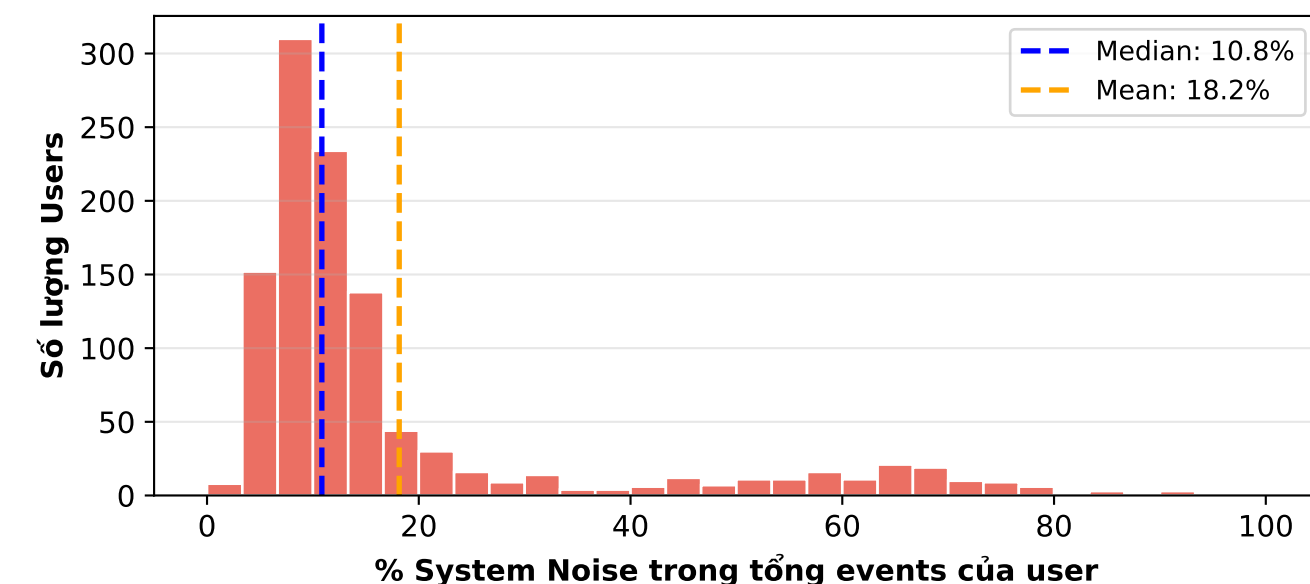
📊 Top 15 Event Types  
(Đỏ = System Noise, Xanh = User Learning)



⚠️ 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



### ⚠️ TÁC ĐỘNG CỦA SYSTEM NOISE

#### 📊 Thống kê:

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

#### ⚠️ Vấn đề:

1. Làm sai lệch chủ sở "hoạt động" của học viên
2. User có nhiều system events ≠ học chăm chỉ
3. Ảnh hưởng clustering: gom nhóm nhóm
4. Ảnh hưởng recommendation: đề xuất sai

#### 🛠️ Giải pháp:

1. Loại bỏ system events trước khi phân tích
2. Chỉ dùng learning-related events
3. Tạo feature "learning\_intensity" riêng
4. Weighted events (learning events > system)