1. **Se quisermos expandir o monitoramento para todos os modelos atualmente em produção, você acha que pode dar algum problema caso haja muitas requisições simultâneas ao mesmo endpoint da API? O que podemos fazer nesse caso?**

   Acredito que possa dar problema sim. Cada requisição feita a um determinado endpoint é tratada como um novo processo a ser computado do lado do servidor. Nesse sentido, podemos entender esse único endpoint como sendo o único processo responsável por devolver a resposta necessária ao usuário. Assim, caso haja muitas requisições, o servidor pode pecar em processar todas elas em tempo aceitável e acabar gerando uma grande fila de requisições a serem tratadas. No entanto, existe uma técnica muito utilizada pelas grandes empresas hoje em dia: **load balancing**. O balanceamento de carga é responsável por controlar e redirecionar o tráfego de rede entre servidores e seus clientes. Para utilizá-lo, podíamos ter vários servidores conectados a nossa aplicação e o load balancer conseguiria fazer o direcionamento adequado das requisições para aqueles servidores que se encontram disponíveis naquele instante. Isso acaba aumentando a disponibilidade da aplicação, bem como sua escalabilidade ao permitir múltiplos acessos ao mesmo endpoint. Podemos entender o objetivo do balanceador como sendo o mesmo de um escalonador de processos. Muitos dos algoritmos de balanceamento, como o método round-robin, são também utilizados em escalonadores de sistemas operacionais.

2. **Que outro problema um modelo de machine learning pode enfrentar em produção que você ache interessante monitorar?**

   Com cada vez mais volume de dados sendo coletados e alimentados no modelo em uso, é importante também monitorar a integridade dos dados a fim de garantir que estes possam seguir pela pipeline sem quebrá-la. Ligado a isso, o monitoramento de outliers é importante de ser feito para justamente permitir reavaliar a estatística utilizada no pré-procesamento com o objetivo de adequar o modelo em produção a nova tendência dos dados. Ademais, o fenômeno do *data drifting* é de certa forma um problema real que ameaça muitas vezes a longevidade do modelo. Cada vez mais novos dados (antes nunca vistos pelo modelo) surgem e são necessários de serem enquadrados na produção para que tal modelo tome as decisões corretas adaptadas a novos grupos de dados que possam vir a surgir.