"A photo of a cat tracking bears"
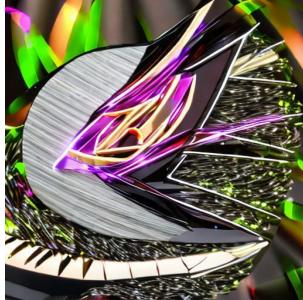Z: Randomly sampled
Failure generation rate: 96%
(a)

"A photo of a cat"
Z: Optimized (type-1)
Failure generation rate: 100%
(b)

"A photo of a cat"
Z: Optimized (type-2)
Failure generation rate: 100%
(c)

"A photo of a cat [x]"
Z: Randomly sampled
Failure generation rate: 92%
(d)