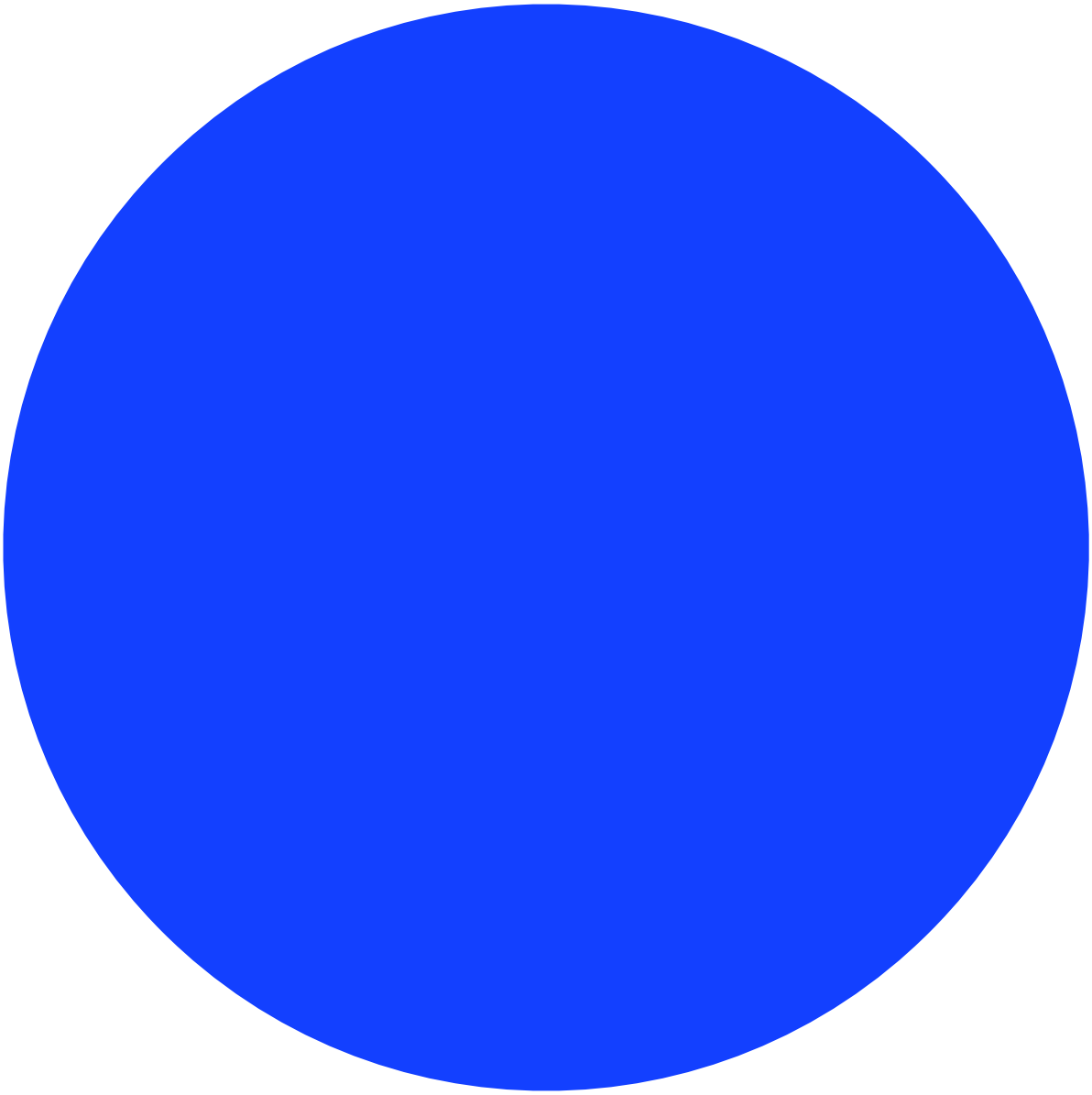
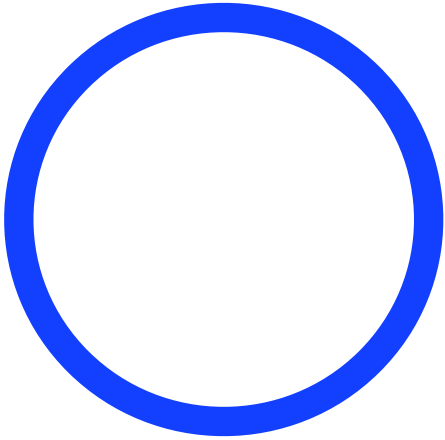


EBU

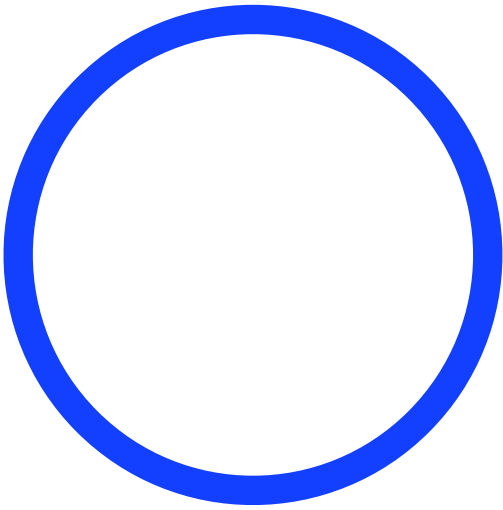
European Broadcasting Union





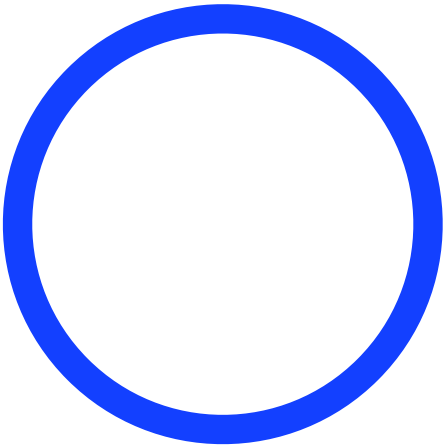
Typical

worflow



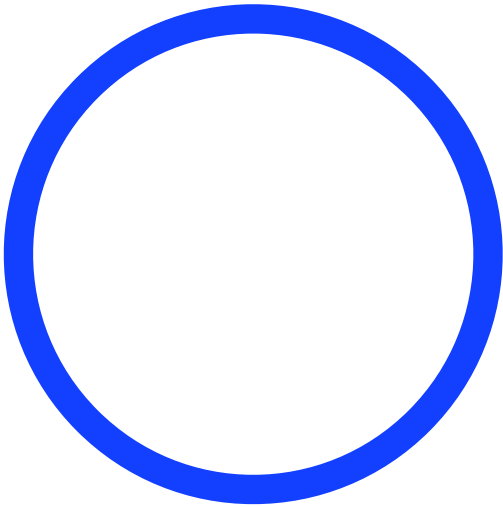
AB Testing -

important!



now is

approrach



Mattnatics!













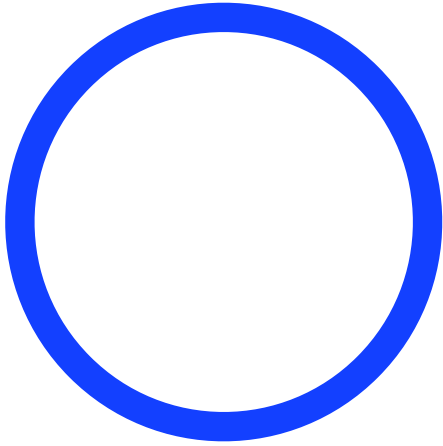












MetriCS

• **Proper evaluation is difficult**

Using just one metric can give us a limited view of how the system works. We

• **Not clear what to measure, how**

• **Things we care about are hard to measure**

- **Many choices that can influence results**

Not a replacement for A/B testing, just more confident before it

should always try to evaluate with different models

Discrepancy model that performs poorly during offline testing

Concisions







